

MECHANISTIC INSIGHTS INTO EDITING SITE SPECIFICITY
OF ADENOSINE DEAMINASES THAT ACT ON RNA:
CRITICAL ROLE OF A CONSERVED LOOP

by

Ashani Kuttan

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biochemistry

The University of Utah

December 2012

Copyright © Ashani Kuttan 2012

All Rights Reserved

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of Ashani Kuttan
has been approved by the following supervisory committee members:

<u>Brenda L. Bass</u>	, Chair	<u>7/11/2012</u> Date Approved
<u>Christopher P. Hill</u>	, Member	<u>7/12/2012</u> Date Approved
<u>David Stillman</u>	, Member	<u>7/11/2012</u> Date Approved
<u>Darrell R. Davis</u>	, Member	<u>7/11/2012</u> Date Approved
<u>Tim Formosa</u>	, Member	<u>7/11/2012</u> Date Approved

and by Christopher P. Hill, Chair of
the Department of Biochemistry

and by Charles A. Wight, Dean of The Graduate School.

ABSTRACT

Adenosine deaminases that act on RNA (ADARs) deaminate adenosines in double-stranded RNA (dsRNA) to produce inosines. The extent an adenosine is edited depends on the sequence context of the target adenosine. Human ADAR2 (hADAR2) has a 5' nearest neighbor preference of U>A>C>G and a 3' preference of G>C>U≈A, but it is not known which amino acids mediate these preferences. Previous studies show that preferences are derived mainly from the catalytic domain. Thus, we adapted a previously reported screen in yeast to identify mutations in the hADAR2 catalytic domain that allow editing of an adenosine in context of a disfavored triplet, GAC. A favored triplet, UAG, was used as the positive control. Hairpin substrates containing disfavored GAC and favored UAG were based on the R/G editing site of GRIA2 (glutamate receptor, ionotropic, AMPA 2) pre-mRNA, a well-studied endogenous substrate for hADAR2.

Four mutants that edited GAC more than WT hADAR2 (E488Q, V493T, N597K and N613K) and one mutant that did not edit GAC (T490A) were further characterized by determining their binding affinity, catalytic rate, base-flipping and preferences to understand the effect of these mutations on ADAR reactivity. Gel-shift assays showed two mutants, N597K and N613K, had ~2-fold higher binding affinity compared to WT hADAR2, suggesting these mutants may have been selected in the screen due to tighter binding. Other mutants E488Q, T490A and V493T, which are on a highly conserved loop

close to the active site, showed similar binding affinity as WT hADAR2 for both UAG and GAC, indicating discrimination was not derived from differences in binding affinity.

We also determined catalytic rates, and probed base-flipping by substituting the target adenosine with the fluorescent base analog 2-aminopurine (2-AP). Remarkably, with both UAG and GAC substrates, mutants with similar binding affinity showed a correlation between catalytic rate and base-flipping, as indicated by a change in 2-AP fluorescence intensity (FI). Our data provide the first information on the residues important for preferences, and point to a conserved loop as key. Unexpectedly, our data suggest that hADAR2's preferences are derived from differences in base-flipping, rather than direct recognition of the neighboring base.

TABLE OF CONTENTS

ABSTRACT.....	iii
LIST OF TABLES.....	vii
LIST OF FIGURES.....	viii
ACKNOWLEDGMENTS.....	x
CHAPTER	
1. INTRODUCTION.....	1
RNA editing.....	1
ADARs and their functional domains.....	2
ADAR evolution.....	7
ADAR editing sites.....	7
ADARs are essential in mammals.....	9
Involvement of ADARs in diseases.....	10
ADAR specificity.....	11
Crystal structure of the hADAR2 catalytic domain.....	14
Mechanism of adenosine deamination by ADARs.....	17
Mutational analysis of residues close to the active site.....	18
Experimental goals.....	21
References.....	21
2. MECHANISTIC INSIGHT INTO ADENOSINE TO INOSINE EDITING SPECIFICITY.....	28
Introduction.....	29
Results.....	31
Discussion.....	53
Materials and methods.....	60
Acknowledgments.....	64
References	65
Supplementary information.....	69

3. PERSPECTIVES.....	81
Summary.....	81
Residues in ADAR involved in base-flipping and preferences.....	81
How are editing sites specifically selected by other deaminases?.....	89
Are ADARs' preferences derived from differential base-flipping?.....	92
What is the rate-limiting step of the deamination reaction?.....	93
Conclusion.....	94
References.....	94
APPENDICES	
A. PURIFICATION OF MUTANT PROTEINS.....	97
B. HUMAN ADAR2 CONTAINING THE Y668H MUTATION	98
C. EQUILIBRIUM EXPERIMENTS.....	100
D. INCUBATION TIMES REQUIRED FOR ACHIEVING ~20% EDITING OF A 418 bp dsRNA.....	102
E. PREFERENCE ASSAYS OF A589V, G336D AND THE DOUBLE MUTANT E488Q/T490A.....	103
F. SEQUENCE OF 418 bp dsRNA USED FOR PREFERENCE ASSAYS.....	106

LIST OF TABLES

<u>Table</u>	<u>Page</u>
2.1 Mutational analysis of mutants identified from the screen.....	36
2.2 Characterization of hADAR2 WT and mutants.....	39
2.S1 Primer sequences.....	79
D.1 Incubation time for achieving ~20 % editing.....	102
E.1 Relative nearest neighbor specificity, S_{rel} , for mutants compared to WT hADAR2	104

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1.1 ADARs deaminate adenosines to create inosines within double stranded regions of RNA.....	3
1.2 Domains of ADAR.....	5
1.3 Crystal structure of hADAR2 catalytic domain.....	15
1.4 Crystal structures of deaminases showing the conserved deaminase motif.....	16
1.5 Proposed mechanism for adenosine to inosine conversion by ADAR.....	19
2.1 Mutants that edit the disfavored <u>GAC</u> hairpin were identified from a screen in yeast.....	32
2.2 Binding affinity is similar to that of wildtype hADAR2 for some mutants, but increased for others.....	38
2.3 Deamination rates for some mutant enzymes are similar to that of WT hADAR2, but others differ.....	42
2.4 2-AP fluorescence assays suggest certain mutants alter base-flipping.....	45
2.5 hADAR2 mutations affect specificity.....	51
2.6 Model describing preference for adenosine in context of <u>UAG</u> compared to <u>GAC</u> , and the role of residues in the conserved loop.....	55
2.S1 Binding affinity of truncN597K/N613K is ~4-fold higher than truncWT.....	74
2.S2 Experiments to test single turnover kinetics.....	75
2.S3 Confirmation that enzymes retain activity for the course of the experiment....	76
2.S4 Control experiments with UA _{2AP} G-28 to determine saturating protein concentrations for the 2-AP fluorescence assay.....	77

2.S5	<i>In vivo</i> α -galactosidase assay to test editing of UAG with R481A mutant.....	78
3.1	Crystal structure of M. Hha1.....	83
3.2	Loops involved in preferences.....	88
3.3	Crystal structures of deaminases showing loops involved in specificity.....	90
A.1	Coomassie blue stained SDS-PAGE analysis of E488Q.....	97
C.1	Equilibrium experiments.....	101
E.1	Preference assays.....	105

ACKNOWLEDGMENTS

I thank my advisor Prof. Brenda L. Bass for giving me the opportunity to work on the ADAR project in her lab, and for her guidance and patience. I am also grateful to her for giving me experimental freedom and being very supportive. I also thank Prof. David Stillman for essential advice and guidance without which the yeast based screen would not have been possible. I would also like to thank all my committee members for advice. I am also grateful to all past and present members of the Bass Lab for scientific discussions and being great resources from whom I have learnt a lot. I would also like to thank all faculty, staff and colleagues in the biochemistry department for being available for advice. Finally, I would like to thank my family for their constant support and encouragement.

CHAPTER 1

INTRODUCTION

RNA editing

RNA editing refers to a post-transcriptional process where the RNA sequence is altered from that encoded by the DNA. RNA editing was first reported in trypanosome mitochondria (1), and shortly after RNA editing was also reported in mammals. The three major kinds of RNA editing are insertion, deletion and modification of nucleotides. Over 100 different post-transcriptional modifications have been observed in RNA of which pseudouridylation is the most abundant (2, 3). The predominant type of RNA editing in mammalian mRNA involves deamination of adenosine to inosine (A-to-I) mediated by ADARs (Adenosine deaminases that act on RNA) (4). Another type of mRNA editing resulting in nucleotide substitution involves deamination of a cytidine to uridine (C-to-U) mediated by APOBEC1 (apolipoprotein B editing catalytic subunit 1). A-to-I RNA editing has also been observed in tRNAs, mediated by ADATs (Adenosine deaminases that act on tRNAs).

Some RNA editing events increase the diversity of proteins

Several mechanisms are used by organisms to produce functionally different proteins from a single gene, thereby increasing the protein diversity. One such

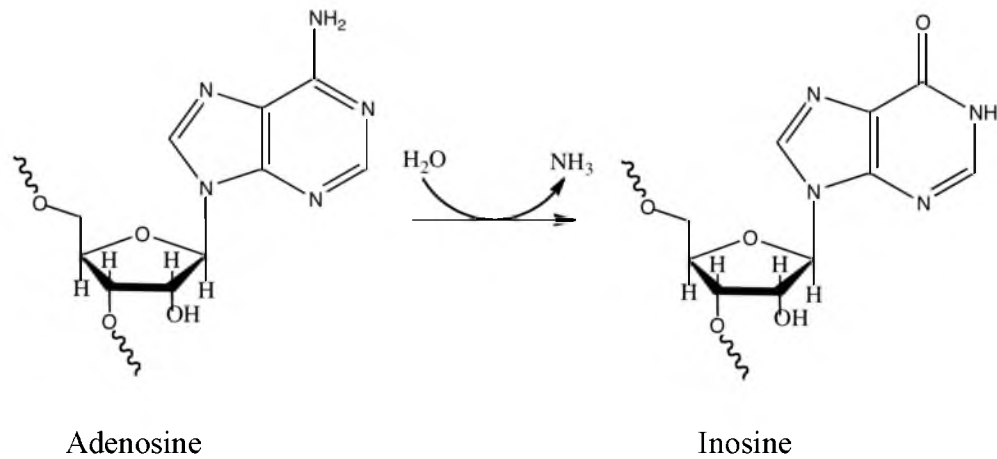
mechanism is the post-transcriptional editing of mRNA. Others include alternative splicing, use of alternative promoters, translational frameshifting, and post-translational modifications. One well-studied example of editing creating a functionally different protein is the C-to-U editing of apolipoprotein B (apoB) mRNA by APOBEC1, a cytidine deaminase that acts on RNA (CDAR). Tissue specific editing of this mRNA in the small intestine converts a glutamine codon into a stop codon resulting in the formation of a truncated apoB 48 instead of the longer apoB 100 (5). The full-length apoB 100 and the truncated apoB 48 have different functions in lipoprotein metabolism. Other well-studied examples are A-to-I editing within the pre-mRNAs of GRIA (glutamate receptor, ionotropic, AMPA 2) and 5-HT_{2C} serotonin receptor (6).

ADARs and their functional domains

ADARs belong to a family of deaminases that include ADA (adenosine deaminases), CDA (cytidine deaminases), ADATs and APOBECs. ADARs convert an adenosine into an inosine (Fig. 1.1) in double-stranded regions of pre-mRNAs, non-coding RNAs and viral RNAs (7-9). Editing in coding regions results in formation of multiple protein isoforms due to recognition of inosine as guanosine by translational and splicing machinery. Additionally, inosine base pairs with a cytidine, thus an A•U base-pair becomes an I•U mismatch and an A•C mismatch becomes an I•C base-pair. Hence, A-to-I editing changes the base-pairing potential of the RNA, thereby altering the RNA structure. Editing also modulates miRNA biogenesis and targeting (7, 10).

ADARs were initially identified in *Xenopus laevis* and have been subsequently characterized in multiple metazoans, including humans, flies, worms and squid, but are

A



B

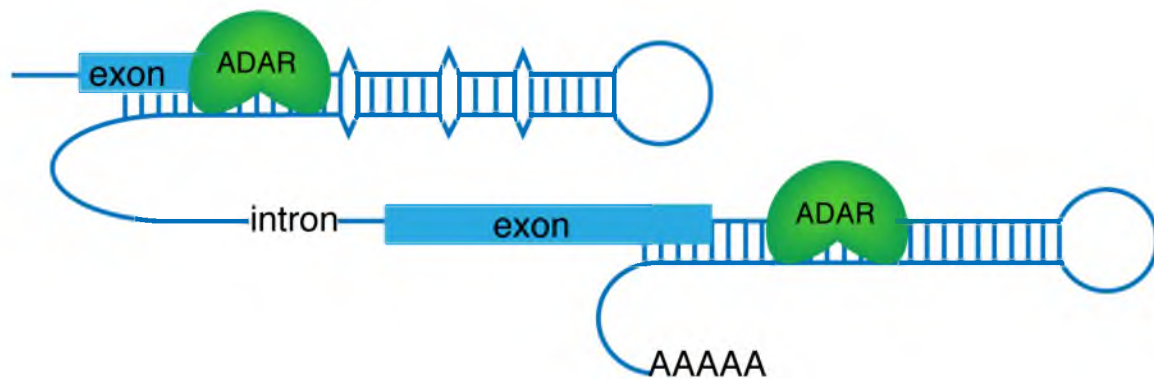


Figure 1.1. **ADARs deaminate adenosines to create inosines within double stranded regions of RNA.** (A) Hydrolytic deamination of adenosine to inosine with the subsequent release of ammonia. (B) A cartoon depicting ADARs targeting double-stranded regions of RNA.

absent in plants and fungi (7, 9). Usually more than one ADAR is found in an organism. In mammals, there are three ADAR genes: ADAR1, ADAR2 and ADAR3, and each has 2-3 double-stranded RNA binding motifs (dsRBMs) on the N-terminus, and a highly conserved deaminase domain on the C-terminus (4, 7) (Fig. 1.2). ADAR1 and ADAR2 are active deaminases, however, no deaminase activity has been observed for ADAR3 despite the presence of all functional domains (11, 12). ADAR1 and ADAR2 are expressed in most tissues, but ADAR3 is expressed exclusively in the brain.

There are two isoforms of ADAR1, generated from alternative promoters: a full-length isoform, ADAR1-L, and a shorter N-terminal truncated version, ADAR1-S (13). ADAR1-L is expressed from an interferon inducible promoter, whereas ADAR1-S is expressed from a constitutive promoter (14). ADAR1-S is initiated from a downstream Met codon that excludes an upstream exon due to alternative splicing. ADAR1-L is found in both the nucleus and the cytoplasm, whereas ADAR1-S localizes mainly to the nucleus (13). Both ADAR1-L and ADAR1-S have an extended amino terminus with Z-DNA binding domains (15), however they differ in that ADAR1-L has two Z-DNA binding domains, $Z\alpha$ and $Z\beta$, while ADAR1-S has only one, $Z\beta$. Both B-DNA and A-RNA can undergo transitions into left-handed helices, Z-DNA and Z-RNA respectively (16). Z-DNA can be formed *in vivo* from negative super-coiling during transcription and *in vitro* under high salt conditions. On the other hand, Z-RNA formation is less favorable energetically than Z-DNA, but evidence suggests Z-RNA is present in both the cytoplasm as well as the nucleolus, although its function is not clear (16). Furthermore, the $Z\alpha$ domain of ADAR1 has been shown to bind Z-RNA suggesting that the *in vivo* function of the $Z\alpha$ domain could be binding Z-RNA (14, 16-18).

A

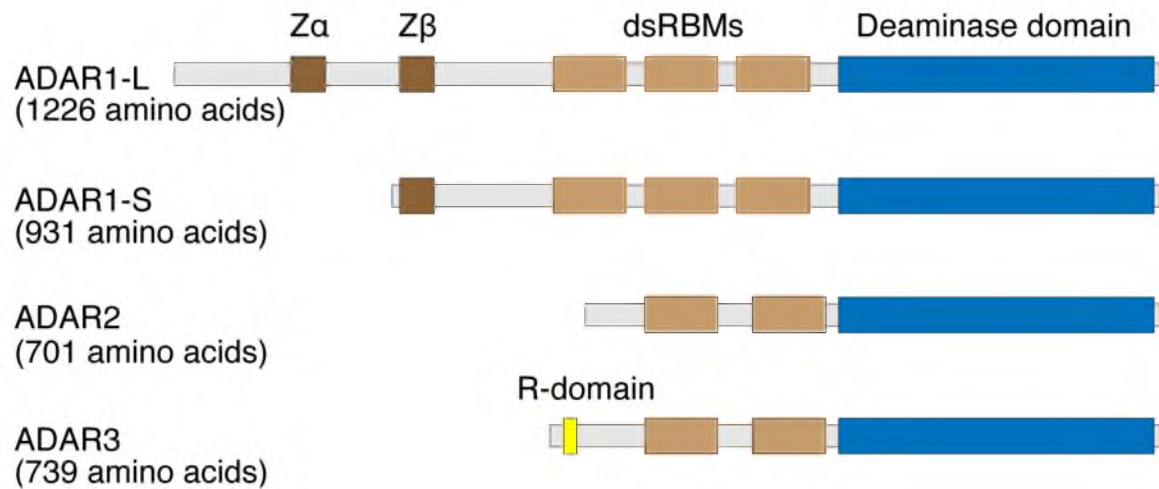


Figure 1.2. **Domains of ADAR.** (A) Mammalian ADARs have a conserved deaminase domain (in blue) at the C-terminus, and either two or three dsRBMs (light brown). ADAR1 has one or two Z-DNA binding motifs on the N-terminus, Z-alpha and Z-beta (dark brown). Shown in yellow is the arginine rich domain. This figure has been adapted from (4).

There are multiple isoforms of ADAR2 produced from alternative splicing events (19-21). A novel ADAR2 splice variant, ADAR2-R, found in vertebrates including humans and mice, has 49 additional amino acids on the N-terminus, produced by incorporating an exon located 18 kilobases upstream of the initially annotated exon one (22). This stretch of 49 additional amino acids includes an arginine-rich domain (R-domain) similar to that observed in ADAR3 where it binds single stranded RNA. ADAR2 expression is regulated by CREB (cyclic AMP response element binding), but the mode of regulation for ADAR3 is not known (4, 23). On studies with human ADARs, no editing activity has been observed with ADAR3 on long synthetic double-stranded RNA (dsRNA), or on known endogenous editing sites like the Q/R, R/G and intronic hotspot1 sites of GRIA2, and the five editing sites of 5-HT_{2C} serotonin receptor (11, 12). However,

ADAR3 reduces the editing efficiency of ADAR1 and ADAR2 *in vitro*, possibly by sequestering the substrates (10).

ADARs have also been studied extensively in *Drosophila melanogaster*, *Caenorhabditis elegans* and squid. *Drosophila melanogaster* has a single ADAR gene, *dADAR*, which is homologous to mammalian ADAR2. *C. elegans* has two ADAR genes, *adr-1* and *adr-2*, of which only *adr-2* encodes an active deaminase. ADARs are not essential in *D. melanogaster* and *C. elegans* (7). Squid has two isoforms of ADAR2, ADAR2a and ADAR2b, of which ADAR2b is active but ADAR2a is inactive.

Other deaminases

ADA and CDA perform A-to-I and C-to-U deaminations, respectively, of free nucleotides. ADATs perform A-to-I deaminations in tRNAs. There are 3 ADATs in eukaryotes; ADAT1, ADAT2 and ADAT3. ADAT1 deaminates A37 of tRNA^{ala} (24), and is not essential. ADAT2/ADAT3 function as a heterodimer and deaminate the wobble position (A34) of multiple tRNAs, which leads to altered codon recognition (25); this post-transcriptional process is a major determinant of genetic code degeneracy. This might explain why ADAT2/ADAT3 are essential. The deaminase domain of ADAT2/ADAT3 is more similar to the deaminase domain of CDA than ADA (26), and the deaminase domain of ADAT1 is similar to that of ADARs. All ADATs lack dsRBMs.

The APOBEC family includes a number of enzymes, and was named after the first family member cloned, APOBEC1, which targets apoB mRNA. APOBEC-1 functions as a homodimer, and as part of a multisubunit complex, to perform editing (25, 27). Further, APOBEC-1 does not have a separate RNA binding motif like ADARs, instead residues in

its active site also mediate RNA binding (28). APOBEC2 is a muscle specific protein, which is catalytically inactive, whereas AID (activation induced deaminase) and APOBEC3 act on ssDNA (29). AID is involved in immunoglobulin class switch recombination as well as in the somatic hypermutation that occurs during an immune response, while APOBEC3 performs C-to-U deaminations in retroviral DNA replication intermediates (25).

ADAR evolution

Phylogenetic analysis suggests that CDA gave rise to ADAT2 and ADAT3, which in turn gave rise to ADAT1, and after acquisition of dsRBMs, ADARs were formed. It has been proposed that APOBEC1 likely evolved separately from CDA (25, 27). Sequence alignments have shown that ADARs are also similar to DNA cytosine-5-methyl transferases and adenine-N6-amino methyl transferases (30). Thus, ADARs and CDAs may be evolutionarily related to methyl transferases, since methyl transferases can target both adenine and cytosine.

ADAR editing sites

A-to-I editing in coding regions

Two of the most studied ADAR editing sites are in GRIA2 pre-mRNA. One is the Q/R site in exon 11, where A-to-I editing recodes the uncharged Gln (Q) into positively charged Arg (R) in the receptor's ion channel, making calcium influx unfavourable. Almost 99% of the GRIA2 subunit in the cell is edited to the R form (4, 7). Under-editing at this site results in increased calcium influx leading to excitotoxicity and

neurodegeneration. Editing at the Q/R site also regulates tetramerization of the AMPA receptors (31). The second editing site is the R/G site in exon 13, where editing leads to formation of a protein isoform with an Arg replaced by a Gly. Editing of the R/G site results in a faster recovery rate of the receptor channel after desensitization (32). The R/G site is edited by both ADAR1 and ADAR2, whereas the Q/R site is edited solely by ADAR2 (33).

Another well-studied example of editing altering coding potential is editing in the 5-HT_{2C} serotonin receptor pre-mRNA (6). There are five editing sites (A-E) within the sequence that encodes amino acids 156-160, and different combinations of editing at these sites can result in a total of 24 different protein isoforms (34). These five editing sites encode three amino acids within the second intracellular loop of the receptor, which is believed to be involved in G-protein coupling. As these sites get increasingly edited, the receptors become progressively less active (34). Sites A and B are edited by ADAR1, site D is edited by ADAR2, and sites C and E are edited by both ADAR1 and ADAR2 (4).

Rat ADAR2 also edits its own transcript to create an alternatively spliced variant that is inactive (35), and hence modulates its own expression.

A-to-I editing in noncoding regions

miRNAs are produced by the sequential cleavage of pri-miRNAs and pre-miRNAs. Pri-miRNAs form long intramolecular stem-loop structures in the nucleus, and are cleaved by Drosha to produce pre-miRNAs, which are 60-70 nt long. Pre-miRNAs are exported to the cytoplasm and further processed by Dicer into mature miRNAs. Since dsRNA is targeted by both ADARs, as well as by components of the miRNA biogenesis

pathway, A-to-I editing modulates the biogenesis and targeting of many miRNAs. For example, *in vitro* studies demonstrate that editing of pri-miR-142 inhibits cleavage by Drosha, and consistent with this observation, expression of mature miR-142 is higher in *ADAR1*^{-/-} and *ADAR2*^{-/-} mice (36). On the other hand, *in vitro* studies show that A-to-I substitution in pri-miR-151, does not inhibit cleavage by Drosha, but the same substitution in pre-miR-151 inhibits cleavage by Dicer (37). Consistent with this observation, all the pre-miR-151 detected in human amygdala is edited, whereas no mature miR-151 is detected. Further, edited miRNAs may also target a different set of mRNAs. For example, studies using an *in vivo* reporter assay have shown that edited and unedited versions of miR-376a target different transcripts (38). Additionally, *in vitro* studies have also shown that A-to-I editing in long dsRNA diminishes its cleavage by Dicer, thereby decreasing siRNA production (39).

ADARs are essential in mammals

Mice with a heterozygous knockout of ADAR1 are viable, but *ADAR1*^{-/-} homozygous embryos die between embryonic day E11.0 and E12.5, and exhibit widespread apoptosis in many tissues (40). Disintegration of the liver structure and severe defects in hematopoiesis were also observed (41). Similarly, mice with a heterozygous knockout of ADAR2 are also viable, but homozygotes die shortly after birth between postnatal days, P0 to P20, and becomes increasingly seizure prone after P12 (42). The *ADAR2*^{-/-} phenotype is rescued by substituting both alleles of *GRIA2* with alleles encoding an edited version of the Q/R site, indicating that GRIA2 pre-mRNA is the most important substrate for ADAR2 for viability (42). On the other hand *ADAR3*^{-/-}

mice are viable, but it is not clear if ADAR1 or ADAR2 compensates for the loss of ADAR3 (8).

Involvement of ADARs in diseases

Aberrant A-to-I editing levels are observed in various diseases. Under-editing of the Q/R site of GRIA2 pre-mRNA has been implicated in amyotrophic lateral sclerosis (ALS) (43, 44). Over-editing of the R/G site of GRIA2 pre-mRNA has been observed in epilepsy patients (45). An increase in editing of 5-HT_{2C} serotonin receptor pre-mRNA is observed in depression patients and suicide victims (43). However, the cause for these aberrant A-to-I editing levels is not understood.

The locus for dyschromatosis symmetrica hereditaria (DSH), a pigmentary genodermatosis, maps to the *ADAR1* gene (46, 47). To date more than 100 mutations have been reported in the *ADAR1* gene of DSH patients, and these include missense mutations, frame-shift mutations and nonsense mutations. The majority of the mutations, at least 47, are missense mutations in the catalytic domain.

ADARs have also been associated with exceptional longevity in humans. Longevity is strongly familial, and the majority of the genes associated with longevity are primarily associated with lipoprotein metabolism and insulin/ IGF-1 signaling (48-50). In 2009, a study of centenarians showed that 18 SNPs in *ADAR2* and *ADAR3* genes are associated with extreme old age (48).

Consequences of A-to-I editing during viral infections

During viral infections, A-to-I editing by ADARs can increase virus growth in certain instances while decreasing it in other instances (14). A-to-I editing of viral RNAs by ADARs can be either site-selective, as observed in hepatitis delta virus (HDV), Kaposi's sarcoma associated herpes virus and Epstein-Barr virus, or may be nonselective hypermutation that leads to persistence of infection as observed in many viruses.

In the well-studied example of HDV, two forms of delta antigen proteins, which are essential for the viral lifecycle, HDAg-S and HDAg-L, are produced from the same open reading frame by the inclusion of a stop codon. A-to-I editing within this stop codon converts the stop codon into a tryptophan codon allowing translation of the longer HDAg-L. The shorter form, HDAg-S, is required for viral RNA replication and the longer form, HDAg-L, is required for packaging of the viral genome and assembling new viral particles.

An example of nonselective editing is observed in the measles virus, where hypermutation in the matrix M gene leads to production of defective M protein that is associated with persistence of infection.

ADAR specificity

ADARs specifically edit certain adenosines over others. There are two determinants of specificity: selectivity and preferences. Selectivity refers to the fraction of sites edited in a dsRNA, and is dependent on the length of the dsRNA and on the presence of bulges, mismatches and internal loops (51). *In vitro* studies show nonselective editing in completely base-paired dsRNA >50 bp, where ~50-60% of

adenosines are edited at reaction completion. However, adenosines in shorter dsRNA are edited more selectively (52, 53). Internal loops convert long dsRNA into shorter stretches of dsRNA, possibly by affecting the binding register of ADARs, thereby resulting in more selective editing (54). Selective editing in coding regions, results in formation of protein isoforms with altered functions. Selective editing is also observed in pri-miRNA and pre-miRNA, which affects their processing and targeting. Non-selective editing is often observed in noncoding regions of mRNAs, like 5' and 3' untranslated regions (UTRs) and introns.

The extent of A-to-I editing at a particular site depends on the sequence context of the target adenosine, and this specificity is referred to as preferences (55). Human ADAR1 (hADAR1) and human ADAR2 (hADAR2) have the same 5' nearest neighbor preference of U>A>C>G, and a 3' nearest neighbor preference of G>C≈A>U and G>C>U≈A, respectively (56). hADAR1 and hADAR2 truncations comprising only the catalytic domain have the same 5' preference as the full-length proteins, and have a similar but distinct 3' preference of G>C>A>U and C≈G≈A>U, respectively (56). Further, when the deaminase domains of ADAR1 and ADAR2 are switched, the substrate specificity of the chimeric protein tracks with the protein from which the deaminase domain is derived (57). These studies suggest that preferences derive mainly from the catalytic domain; however, it is unclear which amino acids in the catalytic domain mediate preferences.

ADARs also show preferences for the base opposite the target adenosine (57, 58). Thus, C is most favored as the base opposite the target adenosine. A•U is also favored, whereas A•A and A•G are disfavored. This preference for the opposite base is also

determined by the catalytic domain (57).

Substrate recognition and specificity in other deaminases

Another enzyme that targets mRNA is APOBEC1, which targets apoB mRNA. ApoB mRNA has a specific sequence motif that is required for editing by APOBEC-1. This sequence motif includes a regulator, spacer and a mooring sequence (27). APOBEC-1 acts on ssRNA, whereas ADAR acts only on dsRNA. Hence, although both APOBEC-1 and ADARs perform deamination reactions on mRNA, they clearly must have different mechanisms for substrate recognition.

Interestingly, some APOBECs also have a 5' neighbor preference, but different enzymes in this family have preference for different 5' neighbors. For example, AID specifically targets cytidine with a 5' purine, whereas APOBEC3 prefers a 5' pyrimidine. Further, different members of the APOBEC3 family have unique 5' neighbor preferences. For example, APOBEC3C targets Y-C dinucleotides, where Y indicates a pyrimidine, whereas APOBEC3F targets T-C dinucleotides and APOBEC3G targets C-C dinucleotides (59).

ADARs are believed to have evolved from ADATs. The crystal structure of TadA, a member of the ADAT2 family in prokaryotes, in complex with a tRNA^{Arg2}, showed that TadA makes specific contacts with three nucleobases in the anticodon loop and also with two flanking nucleobases (60). These nucleobases are splayed, and hence can access the specific recognition pockets on the TadA surface. However, ADARs are more similar to ADAT1 than ADAT2/ADAT3. Although, both ADAT-1 and ADARs perform A-to-I deaminations, and substantial sequence homology is seen in the deaminase domain,

ADAT-1 lacks dsRBMs, and acts on the anticodon loop of tRNA, unlike ADARs, which act on dsRNA. Hence, these enzymes also clearly must have different mechanisms for substrate recognition and specificity.

Both ADA and CDA deaminate free nucleotides, unlike ADARs, which act on dsRNA. Hence, ADARs would not have any similarity to ADA and CDA with respect to their mechanism for nearest neighbor specificity.

Crystal structure of the hADAR2 catalytic domain

The crystal structure of hADAR2 catalytic domain (PDB code: 1ZY7, Fig. 1.3) shows the catalytic zinc ion in a deep pocket within a roughly spherical structure, which is 40 Å in diameter (61). The active site zinc coordinates with two cysteines (C451 and C516), one histidine (H394) and a water molecule, which is presumed to displace ammonia during the deamination reaction. Residue E396 in the active site hydrogen bonds to the zinc coordinated water, and acts as a proton shuttle. Earlier studies had shown that equivalent residues in ADAR1 (H910, C966, C1036 and E912) are essential for its activity (62). The geometry of zinc coordination observed at the active site of ADAR2 is identical to that observed in the active site of CDA, TadA and APOBECs (PDB code: 1CTU, 2B3J and 3E1U respectively) (61, 63-66). These enzymes form a structurally conserved deaminase motif that includes five β -strands and two α -helices (colored blue and numbered as in TadA, Fig. 1.4). The structure of the ADAR2 catalytic domain revealed that ADARs are more similar to CDA than ADA. ADAs also perform a hydrolytic deamination at the C6 position of adenosine, just like ADARs, however, their deaminase motif has a different structural topology comprised of an 8-strand parallel α/β

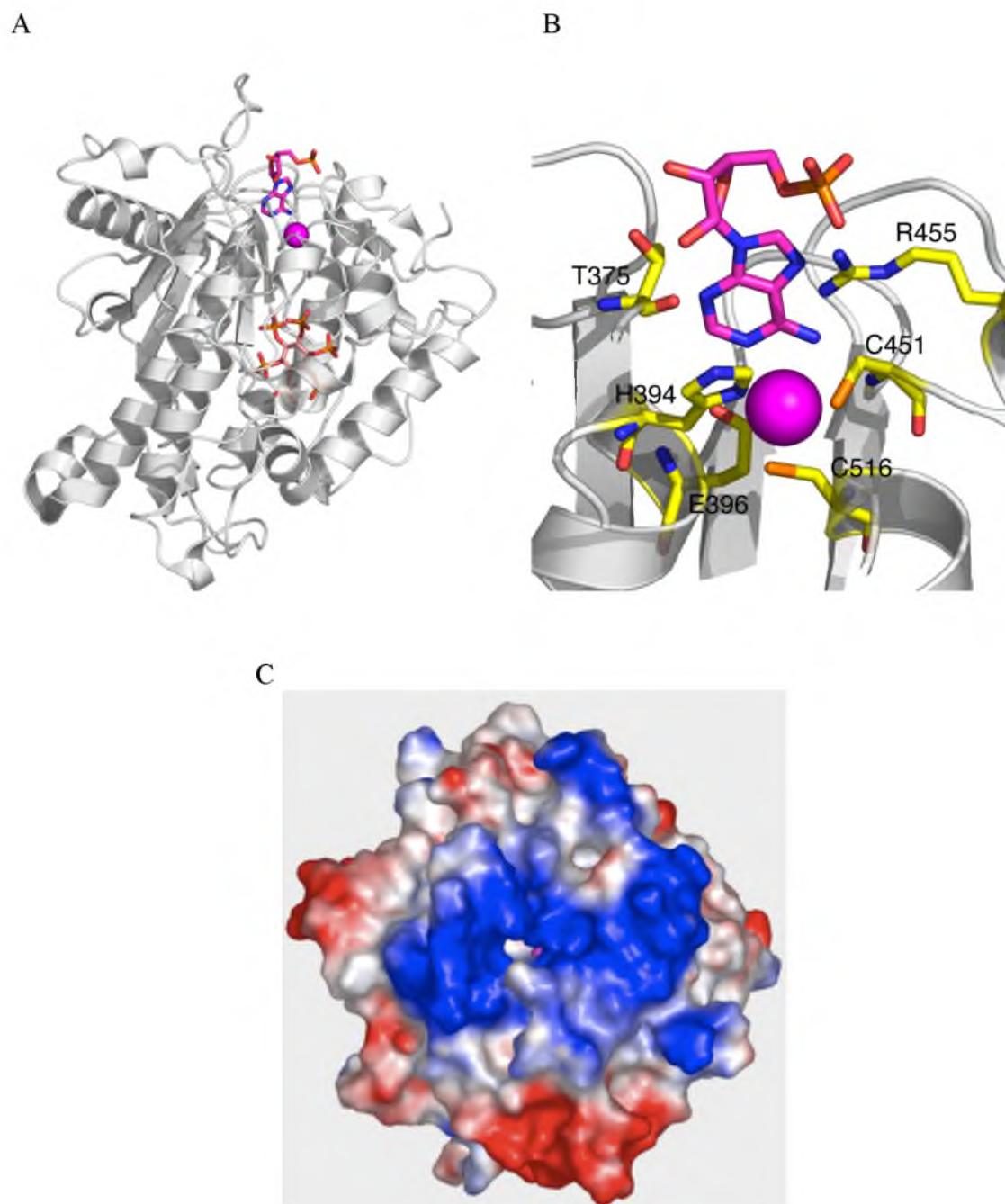


Figure 1.3. **Crystal structure of hADAR2 catalytic domain.** (A) Crystal structure of the catalytic domain of hADAR2 (PDB code: 1ZY7)(61) with Zn (pink sphere), IP6 (orange and red stick), and modeled in AMP (pink stick). (B) A close view of the active site with Zn (pink sphere), modeled in AMP (pink stick), Zn coordinating residues (one histidine and two cysteines), proton shuttle (glutamate) and two residues (T375 and R455) close to the modeled in AMP. (C) Electrostatic surface potential showing a basic patch on the enzyme surface that likely binds dsRNA. The active site Zn (pink sphere) is in a deep pocket on the enzyme surface.

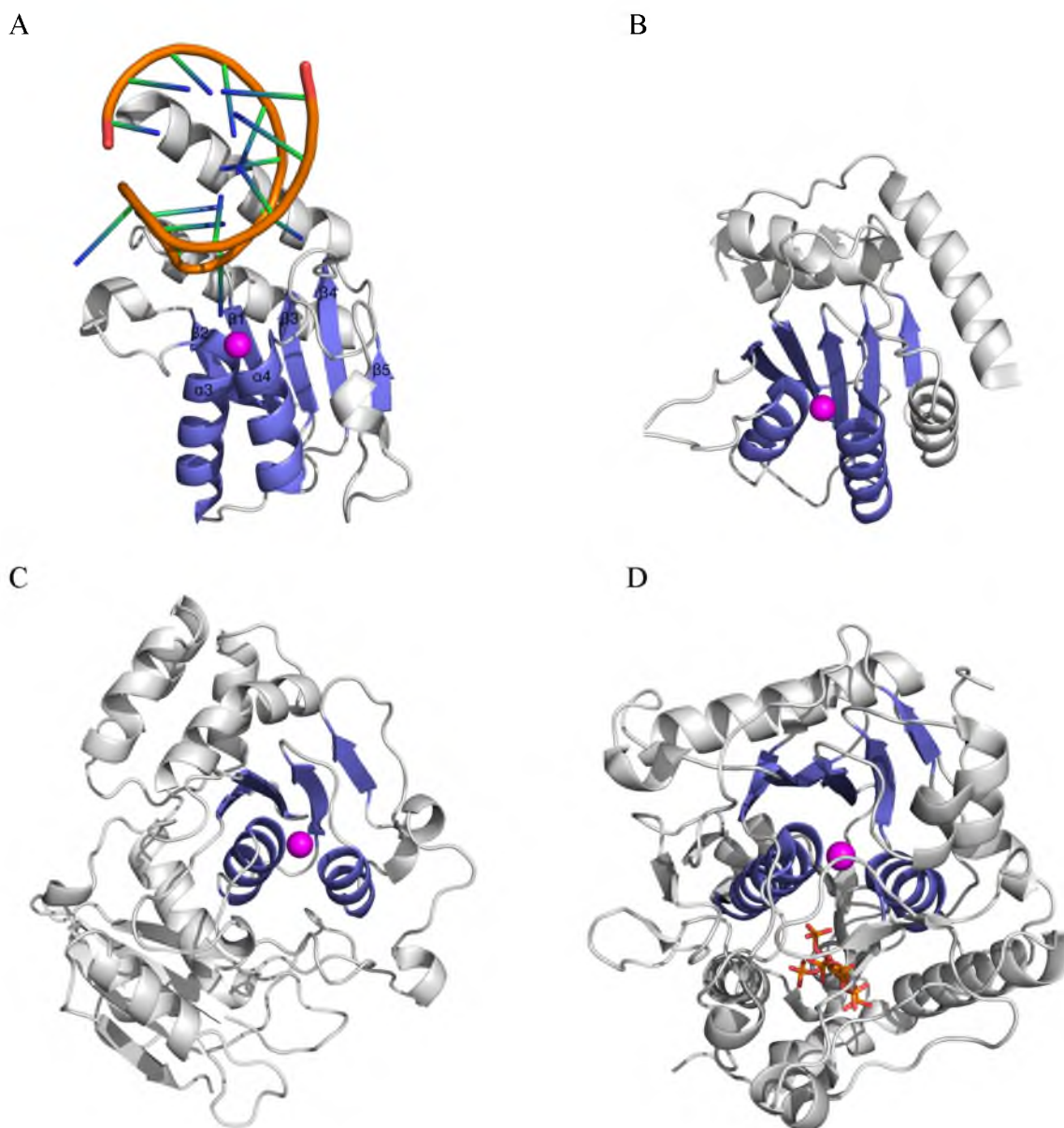


Figure 1.4. Crystal structures of deaminases showing the conserved deaminase motif. (A) The conserved deaminase motif (colored blue) of *S. aureus* TadA in complex with the anticodon stem-loop of tRNA^{Arg2} (PDB code: 2B3J) (60). Five β -strands and two helices are numbered as in TadA. (B, C and D) Conserved deaminase motifs of the APOBEC3G catalytic domain from *E. coli*, CDA from *E. coli*, and the hADAR2 catalytic domain, respectively (colored blue, PDB code: 3E1U, 1CTU, and 1ZY7 respectively) (61, 67, 68). This figure has been adapted from (63).

barrel (63, 64).

C-terminal elements of the deaminase domain of hADAR2 are distinct from those of CDA and TadA, and contribute to the formation of an extremely basic cavity that buries an inositol hexakisphosphate (IP6) molecule and 29 water molecules hydrogen bonded to it. The cavity is lined with basic residues R400, R401, K519, R522, K629, K662, K672 and K690, as well as aromatic residues W523, W687, Y658 and Y668 (61). The majority of the equivalent residues in ADAR1 are conserved. IP6 is presumed to play a structural role by occupying a basic hole in the protein and stabilizing the protein fold. It could also be required for catalysis by fine-tuning the environment of the active site zinc, since IP6 indirectly coordinates with zinc by a relay of hydrogen bonds involving residues K519, D392, K483 and C516 (61). Interestingly, IP6 is also present in ADAT1, although it is absent in ADAT2 and ADAT3, indicating that ADARs are more similar to ADAT1 than ADAT2 or ADAT3. This is consistent with sequence alignments, which indicate that ADARs evolved from ADAT1. Thus, it has been proposed that ADAT1 evolved from ADAT2/ADAT3 by acquiring IP6, and further on acquiring dsRBMs, ADARs were formed (61).

Mechanism of adenosine deamination by ADAR

When deamination reactions were performed on dsRNA, with adenosines uniformly labeled with ^{13}C , the resulting inosine retained all labeled carbons (69). Further, on performing the deamination reaction in ^{18}O labeled water, inosines labeled with ^{18}O at the C6 position was observed (69). This indicates that ADARs use a hydrolytic deamination mechanism to convert adenosine into inosine, similar to ADA and

CDA (70), rather than by a base replacement mechanism, which involves breakage of the glycosidic bond (69).

Based on the mechanism of other enzymes like CDA and TadA that perform hydrolytic deaminations, a similar mechanism has been proposed for ADAR catalysis (Fig. 1.5) (33, 63). Residue E396 in the active site of ADAR abstracts a proton from the zinc-coordinated water, and the resulting hydroxide attacks the C6 position of adenosine. E396 also protonates N1 resulting in the formation of a high energy tetrahedral intermediate, followed by proton transfer from O6 to N6, finally yielding inosine by releasing ammonia. This proposed reaction mechanism for ADAR is consistent with kinetic isotope studies done with an *E. coli* TadA (71).

The adenosine that undergoes nucleophilic attack during the deamination reaction lies deep in the major groove of the dsRNA substrate. Therefore, for this site to be accessible to the enzyme, it is presumed that ADARs utilize a base-flipping mechanism, whereby the adenosines are flipped into the catalytic pocket. Consistent with this mechanism, the active site zinc is in a deep pocket on the enzyme surface, and further, the crystal structure shows a basic patch on the enzyme surface (Fig. 1.3C) that likely binds dsRNA. A base-flipping mechanism for ADAR was supported by studies with 2-aminopurine (2-AP), an adenine analog that is frequently used to probe base-flipping in enzymes (72-74).

Mutational analysis of residues close to the active site

Apart from the zinc coordinating residues and glutamate that are essential for catalytic activity, other residues close to the active site were also identified in the crystal

A

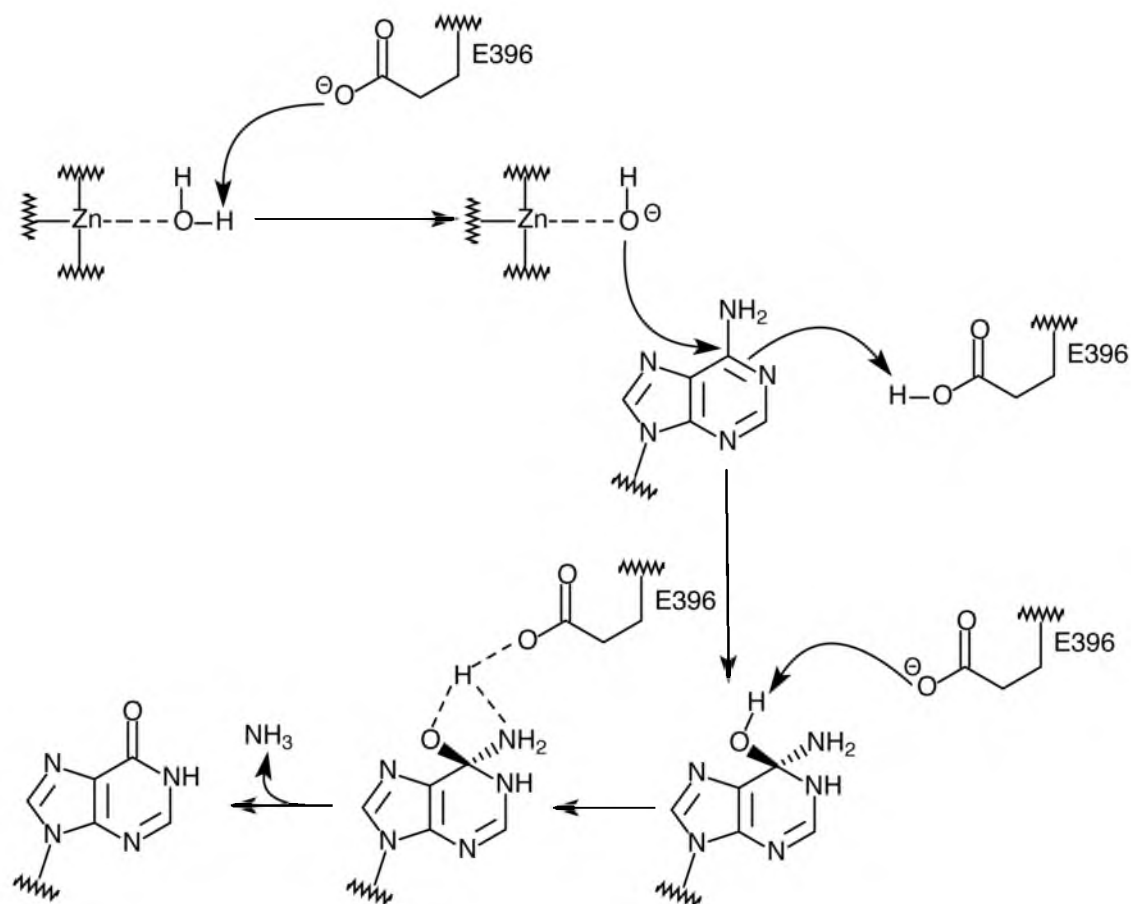


Figure 1.5. **Proposed mechanism for adenosine to inosine conversion by ADAR.**

Residue E396 in the active site of ADAR abstracts a proton from the zinc-coordinated water, and the resulting hydroxide attacks the C6 position of adenosine. E396 further protonates N1 resulting in the formation of a high energy tetrahedral intermediate, followed by proton transfer from O6 to N6, finally yielding inosine by releasing ammonia. This figure has been adapted from (63).

structure of hADAR2 catalytic domain that might have a role in the deamination reaction. On superimposing the crystal structure of the CDA-zebularine complex (zebularine is a cytidine analog and inhibitor of CDA) onto the hADAR2 crystal structure, the ribose of zebularine clashes with a loop of hADAR2 that contains the residue T375 (61). Hence it has been proposed that this loop could be the reason why ADARs do not deaminate cytidine. This clash is absent with the modeled in AMP, since the purine ring of adenosine can access the zinc in the active site with a more shallow penetration than that required by the pyrimidine ring of cytidine. The crystal structure of hADAR2 with the modeled in AMP showed that T375 was in proximity to the 2'-OH of adenosine, and residue K376 could possibly interact with the 3'-phosphodiester of adenosine (75). Modeling in AMP also required a slight repositioning of the R455 side chain, and this side chain is in close proximity to the N7 position of adenosine (61).

The properties of these residues were studied by mutating them into all possible amino acids and performing a yeast-based screen to identify the mutants that had editing activity. The majority of the active mutants of T375 had small hydrophobic residues at this position, although they were less active than the wildtype Thr (75). Large residues at this position were inactive, consistent with its proximity to the 2'-OH of the target adenosine. When both T375 and K376 were simultaneously mutated into all possible amino acids, the majority of the active mutants identified from the screen had a positively charged residue at position 376, but at position 375, only Thr and Cys were observed; both are amino acids that can hydrogen-bond to the 2'-OH of the target adenosine (75). When T375 and R455 were simultaneously mutated, the majority of the active mutants had a Thr, Cys or Ser at position 375, while at position 455, small residues (Gly, Ala, Ser,

Thr) or Arg were observed (76).

Based on these observations, two roles have been proposed for T375: the first is to discriminate the target adenosine from cytidine, and the second is to hydrogen bond with the 2'-OH of the target adenosine (63). This residue is not conserved in ADARs; ADAR1 has an Asn at this position, however, Asn could also possibly form a hydrogen bond with the 2'-OH. These data also suggest that although the side chain of R455 is proximal to the N7 position of adenosine, this residue is not required for activity (76).

Experimental goals

Various studies have shown that the extent an adenosine is edited by ADARs depends on the sequence context of the target adenosine. However, the mechanism underlying nearest neighbor preferences, and the residues in ADAR that confer editing site specificity are major outstanding questions. Answering these questions will be crucial to understand the basis of altered A-to-I editing levels observed in various diseases and to design drugs for specific conditions. In this study, I attempt to answer these questions by performing a screen in yeast to identify mutations in the hADAR2 catalytic domain that will allow editing of an adenosine within a disfavored context, and further characterize these mutants by performing various biochemical experiments to understand the effect of these mutations on ADAR reactivity.

References

1. Benne R, *et al.* (1986) Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* 46(6):819-826.

2. Cantara WA, *et al.* (2011) The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res* 39(Database issue):D195-201.
3. Hamma T & Ferre-D'Amare AR (2006) Pseudouridine synthases. *Chemistry & biology* 13(11):1125-1135.
4. Nishikura K (2006) Editor meets silencer: crosstalk between RNA editing and RNA interference. *Nat Rev Mol Cell Biol* 7(12):919-931.
5. Powell LM, *et al.* (1987) A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *Cell* 50(6):831-840.
6. Burns CM, *et al.* (1997) Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature* 387(6630):303-308.
7. Bass BL (2002) RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem* 71:817-846.
8. Nishikura K (2010) Functions and regulation of RNA editing by ADAR deaminases. *Annu Rev Biochem* 79:321-349.
9. Hundley HA & Bass BL (2010) ADAR editing in double-stranded UTRs and other noncoding RNA sequences. *Trends Biochem Sci* 35(7):377-383.
10. Wulff BE & Nishikura K (2010) Substitutional A-to-I RNA editing. *Wiley Interdiscip Rev RNA* 1(1):90-101.
11. Melcher T, *et al.* (1996) RED2, a brain-specific member of the RNA-specific adenosine deaminase family. *J Biol Chem* 271(50):31795-31798.
12. Chen CX, *et al.* (2000) A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains. *RNA* 6(5):755-767.
13. Patterson JB & Samuel CE (1995) Expression and regulation by interferon of a double-stranded-RNA-specific adenosine deaminase from human cells: evidence for two forms of the deaminase. *Mol Cell Biol* 15(10):5376-5388.
14. Samuel CE (2011) Adenosine deaminases acting on RNA (ADARs) are both antiviral and proviral. *Virology* 411(2):180-193.
15. Herbert A, *et al.* (1997) A Z-DNA binding domain present in the human editing enzyme, double-stranded RNA adenosine deaminase. *Proc Natl Acad Sci U S A* 94(16):8421-8426.

16. Placido D, Brown BA, 2nd, Lowenhaupt K, Rich A, & Athanasiadis A (2007) A left-handed RNA double helix bound by the Z alpha domain of the RNA-editing enzyme ADAR1. *Structure* 15(4):395-404.
17. Brown BA, 2nd, Lowenhaupt K, Wilbert CM, Hanlon EB, & Rich A (2000) The zalpha domain of the editing enzyme dsRNA adenosine deaminase binds left-handed Z-RNA as well as Z-DNA. *Proc Natl Acad Sci U S A* 97(25):13532-13536.
18. Koeris M, Funke L, Shrestha J, Rich A, & Maas S (2005) Modulation of ADAR1 editing activity by Z-RNA in vitro. *Nucleic Acids Res* 33(16):5362-5370.
19. Gerber A, O'Connell MA, & Keller W (1997) Two forms of human double-stranded RNA-specific editase 1 (hRED1) generated by the insertion of an Alu cassette. *RNA* 3(5):453-463.
20. Lai F, Chen CX, Carter KC, & Nishikura K (1997) Editing of glutamate receptor B subunit ion channel RNAs by four alternatively spliced DRADA2 double-stranded RNA adenosine deaminases. *Mol Cell Biol* 17(5):2413-2424.
21. Mittaz L, *et al.* (1997) Cloning of a human RNA editing deaminase (ADARB1) of glutamate receptors that maps to chromosome 21q22.3. *Genomics* 41(2):210-217.
22. Maas S & Gommans WM (2009) Novel exon of mammalian ADAR2 extends open reading frame. *PLoS One* 4(1):e4225.
23. Peng PL, *et al.* (2006) ADAR2-dependent RNA editing of AMPA receptor subunit GluR2 determines vulnerability of neurons in forebrain ischemia. *Neuron* 49(5):719-733.
24. Gerber A, Grosjean H, Melcher T, & Keller W (1998) Tad1p, a yeast tRNA-specific adenosine deaminase, is related to the mammalian pre-mRNA editing enzymes ADAR1 and ADAR2. *EMBO J* 17(16):4780-4789.
25. Gerber AP & Keller W (2001) RNA editing by base deamination: more enzymes, more targets, new mysteries. *Trends Biochem Sci* 26(6):376-384.
26. Gerber AP & Keller W (1999) An adenosine deaminase that generates inosine at the wobble position of tRNAs. *Science* 286(5442):1146-1149.
27. Maas S & Rich A (2000) Changing genetic information through RNA editing. *BioEssays : news and reviews in molecular, cellular and developmental biology* 22(9):790-802.
28. Navaratnam N, *et al.* (1995) Evolutionary origins of apoB mRNA editing: catalysis by a cytidine deaminase that has acquired a novel RNA-binding motif at its active site. *Cell* 81(2):187-195.

29. Conticello SG, Langlois MA, & Neuberger MS (2007) Insights into DNA deaminases. *Nat Struct Mol Biol* 14(1):7-9.
30. Hough RF & Bass BL (1997) Analysis of *Xenopus* dsRNA adenosine deaminase cDNAs reveals similarities to DNA methyltransferases. *RNA* 3(4):356-370.
31. Greger IH, Khatri L, Kong X, & Ziff EB (2003) AMPA receptor tetramerization is mediated by Q/R editing. *Neuron* 40(4):763-774.
32. Lomeli H, *et al.* (1994) Control of kinetic properties of AMPA receptor channels by nuclear RNA editing. *Science* 266(5191):1709-1713.
33. Maydanovich O & Beal PA (2006) Breaking the central dogma by RNA editing. *Chem Rev* 106(8):3397-3411.
34. Gardiner K & Du Y (2006) A-to-I editing of the 5HT_{2C} receptor and behaviour. *Briefings in functional genomics & proteomics* 5(1):37-42.
35. Rueter SM, Dawson TR, & Emeson RB (1999) Regulation of alternative splicing by RNA editing. *Nature* 399(6731):75-80.
36. Yang W, *et al.* (2006) Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat Struct Mol Biol* 13(1):13-21.
37. Kawahara Y, Zinshteyn B, Chendrimada TP, Shiekhata R, & Nishikura K (2007) RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex. *EMBO Rep* 8(8):763-769.
38. Kawahara Y, *et al.* (2007) Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science* 315(5815):1137-1140.
39. Scadden AD & Smith CW (2001) RNAi is antagonized by A→I hyper-editing. *EMBO Rep* 2(12):1107-1111.
40. Wang Q, *et al.* (2004) Stress-induced apoptosis associated with null mutation of ADAR1 RNA editing deaminase gene. *J Biol Chem* 279(6):4952-4961.
41. Hartner JC, *et al.* (2004) Liver disintegration in the mouse embryo caused by deficiency in the RNA-editing enzyme ADAR1. *J Biol Chem* 279(6):4894-4902.
42. Higuchi M, *et al.* (2000) Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. *Nature* 406(6791):78-81.
43. Maas S, Kawahara Y, Tamburro KM, & Nishikura K (2006) A-to-I RNA editing and human disease. *RNA Biol* 3(1):1-9.

44. Kwak S & Kawahara Y (2005) Deficient RNA editing of GluR2 and neuronal death in amyotrophic lateral sclerosis. *J Mol Med (Berl)* 83(2):110-120.
45. Vollmar W, *et al.* (2004) RNA editing (R/G site) and flip-flop splicing of the AMPA receptor subunit GluR2 in nervous tissue of epilepsy patients. *Neurobiol Dis* 15(2):371-379.
46. Zhang XJ, *et al.* (2003) Identification of a locus for dyschromatosis symmetrica hereditaria at chromosome 1q11-1q21. *J Invest Dermatol* 120(5):776-780.
47. Miyamura Y, *et al.* (2003) Mutations of the RNA-specific adenosine deaminase gene (DSRAD) are involved in dyschromatosis symmetrica hereditaria. *Am J Hum Genet* 73(3):693-699.
48. Sebastiani P, *et al.* (2009) RNA editing genes associated with extreme old age in humans and with lifespan in *C. elegans*. *PLoS One* 4(12):e8210.
49. Willcox BJ, *et al.* (2008) FOXO3A genotype is strongly associated with human longevity. *Proc Natl Acad Sci U S A* 105(37):13987-13992.
50. Flachsbart F, *et al.* (2009) Association of FOXO3A variation with human longevity confirmed in German centenarians. *Proc Natl Acad Sci U S A* 106(8):2700-2705.
51. Bass BL (1997) RNA editing and hypermutation by adenosine deamination. *Trends Biochem Sci* 22(5):157-162.
52. Polson AG & Bass BL (1994) Preferential selection of adenosines for modification by double-stranded RNA adenosine deaminase. *EMBO J* 13(23):5701-5711.
53. Nishikura K, *et al.* (1991) Substrate specificity of the dsRNA unwinding/modifying activity. *EMBO J* 10(11):3523-3532.
54. Lehmann KA & Bass BL (1999) The importance of internal loops within RNA substrates of ADAR1. *J Mol Biol* 291(1):1-13.
55. Lehmann KA & Bass BL (2000) Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. *Biochemistry* 39(42):12875-12884.
56. Eggington JM, Greene T, & Bass BL (2011) Predicting sites of ADAR editing in double-stranded RNA. *Nat Commun* 2:319.
57. Wong SK, Sato S, & Lazinski DW (2001) Substrate recognition by ADAR1 and ADAR2. *RNA* 7(6):846-858.

58. Kallman AM, Sahlin M, & Ohman M (2003) ADAR2 A-->I editing: site selectivity and editing efficiency are separate events. *Nucleic Acids Res* 31(16):4874-4881.
59. Langlois MA, Beale RC, Conticello SG, & Neuberger MS (2005) Mutational comparison of the single-domained APOBEC3C and double-domained APOBEC3F/G anti-retroviral cytidine deaminases provides insight into their DNA target site specificities. *Nucleic Acids Res* 33(6):1913-1923.
60. Losey HC, Ruthenburg AJ, & Verdine GL (2006) Crystal structure of *Staphylococcus aureus* tRNA adenosine deaminase TadA in complex with RNA. *Nat Struct Mol Biol* 13(2):153-159.
61. Macbeth MR, *et al.* (2005) Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science* 309(5740):1534-1539.
62. Lai F, Drakas R, & Nishikura K (1995) Mutagenic analysis of double-stranded RNA adenosine deaminase, a candidate enzyme for RNA editing of glutamate-gated ion channel transcripts. *J Biol Chem* 270(29):17098-17105.
63. Goodman RA, Macbeth MR, & Beal PA (2012) ADAR proteins: structure and catalytic mechanism. *Curr Top Microbiol Immunol* 353:1-33.
64. Wilson DK, Rudolph FB, & Quijcho FA (1991) Atomic structure of adenosine deaminase complexed with a transition-state analog: understanding catalysis and immunodeficiency mutations. *Science* 252(5010):1278-1284.
65. Betts L, Xiang S, Short SA, Wolfenden R, & Carter CW, Jr. (1994) Cytidine deaminase. The 2.3 Å crystal structure of an enzyme: transition-state analog complex. *J Mol Biol* 235(2):635-656.
66. Kuratani M, *et al.* (2005) Crystal structure of tRNA adenosine deaminase (TadA) from *Aquifex aeolicus*. *J Biol Chem* 280(16):16002-16008.
67. Holden LG, *et al.* (2008) Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature* 456(7218):121-124.
68. Xiang S, Short SA, Wolfenden R, & Carter CW, Jr. (1995) Transition-state selectivity for a single hydroxyl group during catalysis by cytidine deaminase. *Biochemistry* 34(14):4516-4523.
69. Polson AG, Crain PF, Pomerantz SC, McCloskey JA, & Bass BL (1991) The mechanism of adenosine to inosine conversion by the double-stranded RNA unwinding/modifying activity: a high-performance liquid chromatography-mass spectrometry analysis. *Biochemistry* 30(49):11507-11514.

70. Carter CW, Jr. (1995) The nucleoside deaminases for cytidine and adenosine: structure, transition state stabilization, mechanism, and evolution. *Biochimie* 77(1-2):92-98.
71. Luo M & Schramm VL (2008) Transition state structure of E. coli tRNA-specific adenosine deaminase. *J Am Chem Soc* 130(8):2649-2655.
72. Stephens OM, Yi-Brunozzi HY, & Beal PA (2000) Analysis of the RNA-editing reaction of ADAR2 with structural and fluorescent analogues of the GluR-B R/G editing site. *Biochemistry* 39(40):12243-12251.
73. Yi-Brunozzi HY, Stephens OM, & Beal PA (2001) Conformational changes that occur during an RNA-editing adenosine deamination reaction. *J Biol Chem* 276(41):37827-37833.
74. Lenz T, *et al.* (2007) 2-Aminopurine flipped into the active site of the adenine-specific DNA methyltransferase M.TaqI: crystal structures and time-resolved fluorescence. *J Am Chem Soc* 129(19):6240-6248.
75. Pokharel S & Beal PA (2006) High-throughput screening for functional adenosine to inosine RNA editing systems. *ACS Chem Biol* 1(12):761-765.
76. Pokharel S, *et al.* (2009) Matching active site and substrate structures for an RNA editing reaction. *J Am Chem Soc* 131(33):11882-11891.

CHAPTER 2

MECHANISTIC INSIGHT INTO ADENOSINE TO INOSINE EDITING SPECIFICITY*

Adenosine deaminases that act on RNA (ADARs) deaminate adenosines in double-stranded RNA (dsRNA) to produce inosines. ADARs are essential in mammals and are particularly important in the nervous system. Altered levels of adenosine to inosine (A-to-I) editing are observed in several diseases. The extent an adenosine is edited depends on sequence context. Human ADAR2 (hADAR2) has 5' and 3' neighbor preferences, but which amino acids mediate these preferences, and by what mechanism, is unknown. We performed a screen in yeast to identify mutations in the hADAR2 catalytic domain that allow editing of an adenosine within a disfavored triplet. Binding affinity, catalytic rate, base-flipping and preferences were monitored to understand effects of the mutations on ADAR reactivity. Our data provide the first information on the amino acids that affect preferences, and point to a conserved loop as being of key importance. Unexpectedly, our data suggest that hADAR2's preferences derive from differential base-flipping, rather than direct recognition of neighboring bases. Our studies set the stage for understanding the basis of altered editing levels in disease and developing therapeutic reagents.

* Kuttan A & Bass BL (2012) Mechanistic insights into editing-site specificity of ADARs. *Proc Natl Acad Sci U S A* doi: 10.1073/pnas.1212548109

Introduction

ADARs target double-stranded regions of pre-mRNAs, non-coding RNAs, and viral RNAs, deaminating adenosines to create inosines (1-3). Inosine is recognized as guanosine, and thus, A-to-I editing in a pre-mRNA can alter codons and splice-forms, leading to multiple protein isoforms from a single gene. ADARs also alter miRNA and endogenous siRNA biogenesis and targeting (2-5). A-to-I editing of viral RNAs can reduce virus growth as well as enhance it (6).

ADARs are found in most metazoans, and often more than one ADAR exists in an organism. For example, there are three mammalian ADAR genes: ADAR1, ADAR2 and ADAR3, and each has two or three N-terminal dsRNA binding motifs (dsRBMs) and a highly conserved C-terminal deaminase domain. ADAR1 and ADAR2 are active deaminases, but enzymatic activity has not been observed with ADAR3 (2, 7, 8).

Two of the most studied ADAR substrates are the pre-mRNAs of GRIA2 (glutamate receptor, ionotropic, AMPA 2) and 5-HT_{2C} serotonin receptor. GRIA2 pre-mRNA has two editing sites, one that recodes glutamine into arginine (Q/R), and another that recodes an arginine into glycine (R/G). Aberrant A-to-I editing is correlated with several diseases (9). For example, under-editing of the Q/R site of GRIA2 pre-mRNA is implicated in amyotrophic lateral sclerosis (ALS) (9, 10), overediting of the R/G site is observed in epilepsy patients (11), and an increase in editing of 5-HT_{2C} serotonin receptor pre-mRNA is observed in depression patients and suicide victims (9). In addition, the locus for dyschromatosis symmetrica hereditaria (DSH), a pigmentary genodermatosis, maps to the ADAR1 gene (12, 13). The mechanistic basis for altered levels of editing in various diseases is entirely unclear.

The fraction of sites edited in a dsRNA, referred to as selectivity, depends on its length and whether it contains mismatches, bulges and internal loops (3, 14). *In vitro* studies show that nonselective editing occurs in completely base-paired dsRNA ≥ 50 base-pair (bp), whereas adenosines in shorter dsRNA, or that containing mismatches, bulges and loops, are edited more selectively (15-17).

The extent of A-to-I editing at a particular site depends on sequence context, and these rules are referred to as preferences (15, 18). Human ADAR1 (hADAR1) and human ADAR2 (hADAR2) have a 5' nearest neighbor preference of U>A>C>G, and a 3' nearest neighbor preference of G>C \approx A>U and G>C>U \approx A, respectively (19). Truncated forms of hADAR1 and hADAR2 comprising only the catalytic domain have the same 5' preference as the full-length proteins and similar but distinct 3' preferences (G>C>A>U and C \approx G \approx A>U, respectively) (19). Further, when the deaminase domains of hADAR1 and hADAR2 are switched, substrate specificity of the chimeric protein tracks with its deaminase domain (20). These studies suggest that preferences derive mainly from the catalytic domain. However, it is not known which amino acids in the catalytic domain mediate preferences.

To identify the amino acids that mediate preferences, we performed a screen for mutations within the hADAR2 catalytic domain that allow editing of an adenosine in a poor sequence context. Collectively, the hADAR2 variants we identified point to a conserved loop near the active site as important for preferences. Unexpectedly, our data suggest that hADAR2's preferences derive from differential base-flipping, rather than direct recognition of the neighboring bases. These studies offer insight in regard to the altered editing levels correlated with disease and set the stage for developing therapeutic

reagents.

Results

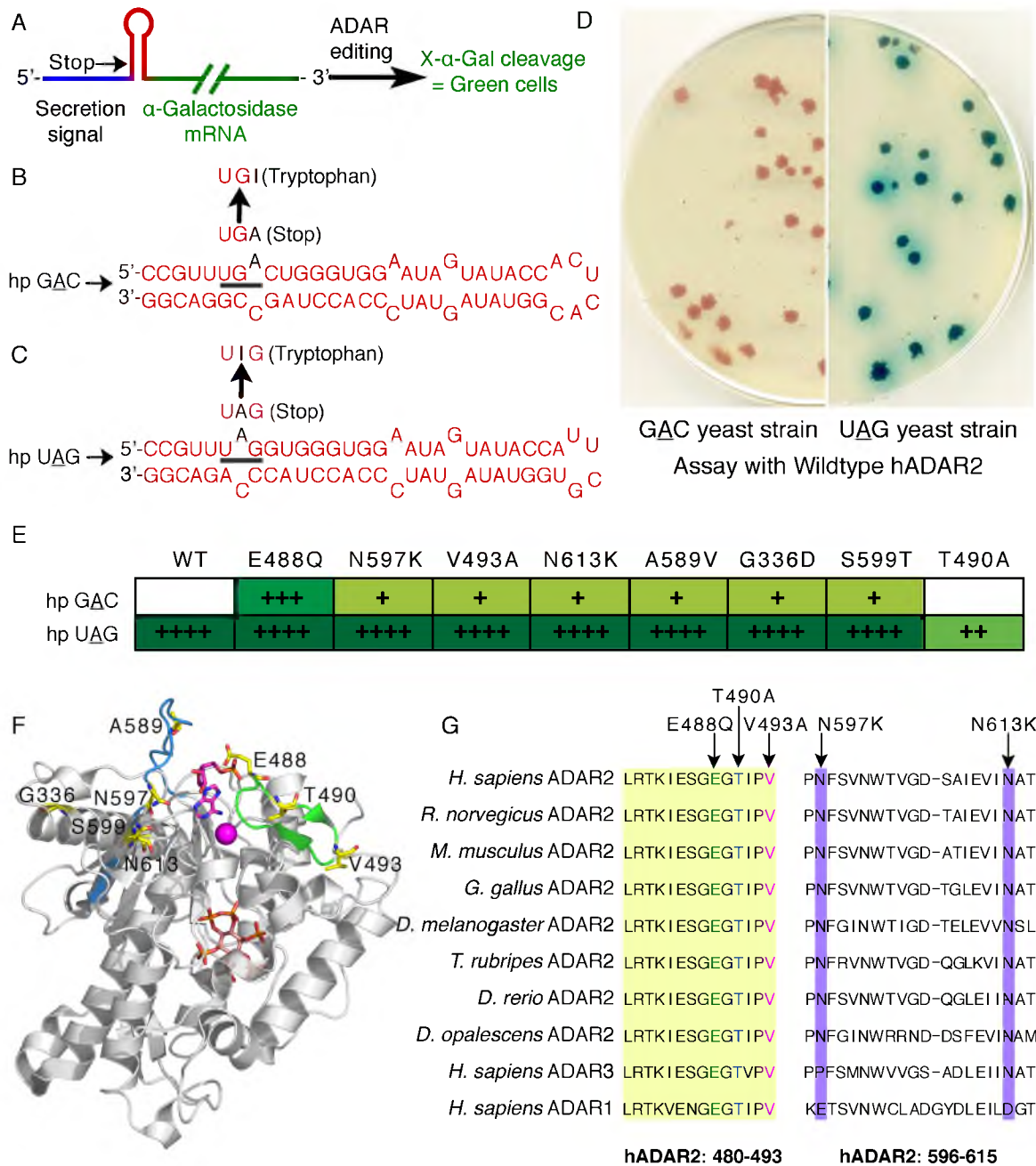
A screen identifies residues in the hADAR2 catalytic domain that affect preferences

For both hADAR1 and hADAR2, nearest neighbor preferences derive mainly from the catalytic domain (19, 20). A crystal structure of the catalytic domain of hADAR2 has been solved (21), and thus, to facilitate our analysis, we focused on this enzyme. We adapted a previously reported screen in *Saccharomyces cerevisiae* (22) to identify mutations in the hADAR2 catalytic domain that allow editing of an adenosine in context of a disfavored triplet, GAC.

The screen relied on a hairpin-reporter that was introduced into the chromosome of the haploid yeast strain, W303 α , under the control of a constitutive promoter, *ADHI*. The hairpin (red, Fig. 2.1A), contained an ADAR editing site within a stop codon either in a disfavored context, UGAC (Fig. 2.1B), or a favored context, UAG (Fig. 2.1C), and its sequence was based on the R/G editing site of GRIA2 pre-mRNA. Editing of either stop codon created a tryptophan codon, allowing expression of the downstream α -galactosidase reporter, and turning yeast colonies green on X- α -gal plates. The sequence upstream of the RNA hairpin (blue, Fig. 2.1A) was the signal sequence for secretion of α -galactosidase.

S. cerevisiae lack an ADAR gene, an essential part of the screen. hADAR2 was introduced in the low copy CEN-vector to ensure uniform protein expression. Expression was under the control of an inducible *GAL* promoter to facilitate induction by galactose

Figure 2.1. Mutants that edit the disfavored GAC hairpin were identified from a screen in yeast. (A) Schematic of the hairpin-reporter used in the screen. The hairpin, in red, is an ADAR substrate with the target adenosine in context of a stop codon that must be edited for expression of the downstream α -galactosidase reporter. (B and C) Sequences of the GAC and UAG hairpins with the target adenosine within a disfavored and favored triplet respectively. (D) Control experiments showing CM-URA plates with yeast colonies that have either a GAC or UAG hairpin-reporter integrated into a chromosome, and transformed with WT hADAR2. (E) Mutants identified from the screen, listed left to right in terms of decreasing green intensity of yeast colonies, an indication of decreasing *in vivo* editing efficiency. Green intensity of yeast colonies with the control UAG hairpin-reporter is also indicated. “++++” indicates that yeast colonies started turning green in ≤ 4 days, “+++” and “++” indicates 5-7 days and “+” indicates low levels of editing taking 2-5 weeks to turn faint green. (F) Mutated residues (yellow sticks) mapped onto the crystal structure of the catalytic domain of hADAR2 (PDB code 1ZY7) (21) with Zn (pink sphere), IP6 (orange and red stick), and modeled in AMP (pink stick). (G) Alignment of hADAR1, hADAR2, hADAR3, and ADAR2 from different species. Mutants that were further characterized are indicated. Highlighted in yellow is the highly conserved loop that includes two beta strands and comprises 14 residues.



after replica plating. Control experiments established that introduction of WT hADAR2 into strains containing the hairpin-reporter with the favored UAG editing site (UAG yeast strain) allowed expression of α -galactosidase, while introduction into the strain containing the hairpin-reporter with the disfavored GAC editing site (GAC yeast strain) did not (Fig. 2.1D).

The hADAR2 catalytic domain was randomly mutagenized by error prone PCR to attain a mutation rate of 0-4 mutations per kilobase, and introduced into the GAC yeast strain in the context of the full-length protein (see Materials and methods). 35,000 colonies were screened, and 24 positives were obtained that could edit GAC more than WT hADAR2. Seven positives had single mutations in the catalytic domain, and E488Q was by far the best in this category (Fig. 2.1E). Four mutations appeared more than once, suggesting the screen was saturated. Plasmids were also rescued from some white colonies, representative of mutants that did not edit GAC, and sequencing verified that these were mostly WT hADAR2 or variants with stop codons in the open reading frame. One of these, T490A, was at an interesting location between two residues mutated in the positives, E488Q and V493A. All mutant forms of hADAR2 were introduced into the UAG yeast strain, and all seven positives retained the ability to edit UAG, whereas T490A edited UAG poorly (Fig. 2.1E). We did not identify any mutants showing a reversed preference, which could edit an adenosine within GAC but not UAG.

When the identified mutations were mapped onto the crystal structure of the hADAR2 catalytic domain (21), most mapped onto the surface of the predicted RNA binding site (21) (Fig. 2.1F). E488Q, T490A and V493A were on a highly conserved loop that includes 2 beta strands and comprises 14 residues (green, Fig. 2.1F; amino acids 480

to 493, shaded yellow, Fig. 2.1G). Other mutations, A589V, N597K and S599T mapped onto another loop and a beta strand on the protein surface (blue, Fig. 2.1F), and N613K mapped nearby. Identification of two proximal asparagine to lysine mutations suggested that these mutants may have been selected because they improved RNA binding. Residues N597 and N613 are conserved in ADAR2 from different species, but are negatively charged residues in ADAR1 (Fig. 2.1G).

The four mutants that showed maximal editing of GAC, as well as T490A, were selected for further characterization. Additional mutagenesis was performed to understand the properties required at these positions. PCR libraries encoding all possible amino acids at each of the 5 residues were created, and introduced into either the GAC yeast strain (E488, V493, N597 and N613; see Materials and methods) or both the GAC and UAG yeast strains (T490). At least 15-30 positives and as many negatives were selected, sequenced and retransformed into the GAC and UAG yeast strains. Most of the negatives had stop codons or frame shifts in the open reading frame.

At position 488, a variety of amino acids were able to substitute for Glu to allow editing of UAG, but only Gln and Asn, polar, uncharged amino acids with an amide side chain, allowed editing of GAC (Table 2.1). Substituting E488 with large hydrophobic residues like Phe, Trp and Leu resulted in loss of editing of both UAG and GAC. We did not identify any amino acids at position 490 that allowed editing at GAC. Further, only Ser and Cys could replace Thr for editing UAG, suggesting that the predicted hydrogen bond (21) from the side chain of T490 to R481 is important. At position 493, a less hydrophobic amino acid, Ala, and polar, uncharged amino acids with a hydroxyl side chain, Thr and Ser, edited GAC more than the WT Val. At position 597 and 613, only

Table 2.1. **Mutational analysis of mutants identified from the screen**

Residue	Amino acid substitution ^a	Editing ^b	
		<u>U</u>A<u>G</u>	<u>G</u>A<u>C</u>
488	Q (25), N (3)	++++	+++
	E (2), A (1), S (1), M (1), R (1)	++++	–
	F (1), L (3), W (1)	–	–
490	T (28), C (8), S (8)	++++	–
	A (3)	++	–
	F (1), Y (1)	+	–
	R (2), K (1), P (3), E (2)	–	–
493	T (4), S (13), A (4)	++++	+
	V (1), R (1), D (1), P (1), G (1)	++++	–
597	K (19), R (13)	++++	+
	N (3), A (1), E (1), H (3), G (1), Y (1)	++++	–
	F (2)	–	–
613	K (6), R (7)	++++	+
	N (1), A (1), E (1)	++++	–

^a Wildtype residues are in bold, and the number beside each amino acid substitution indicates number of clones isolated.

^b Extent of editing of UAG and GAC hairpin-reporters *in vivo* as determined from green intensity of yeast colonies on X- α -Gal plates.

positively charged residues, Lys and Arg, allowed editing of GAC, further supporting the idea that these residues were selected in the screen due to improved RNA binding.

hADAR2 catalytic domain mutants separate into

two classes based on binding affinity

WT and mutant hADAR2 proteins were purified to homogeneity (see Materials and methods) and subjected to *in vitro* characterization. We first performed gel mobility shift assays comparing the proteins for binding to UAG or GAC hairpins that were chemically synthesized and ^{32}P 5' end labeled. Representative gel shifts are shown for WT hADAR2 and the N597K variant (Fig. 2.2A and B). WT hADAR2 and all mutant forms showed the formation of two protein-RNA complexes, as observed in previous studies (23). For all proteins tested, a mobility shift was first observed at a protein concentration of ≤ 1.5 nM, and at high protein concentrations of ~ 50 -100 nM, a second, slower mobility shift appeared, likely due to a second binding event on the RNA. RNA was almost completely bound at a protein concentration of 500 nM. K_d values were determined for the complex represented by the first, fast mobility shift, and binding isotherms are shown in Fig. 2.2C and D.

WT hADAR2 had a K_d of ~ 2.1 nM for both UAG and GAC hairpins, emphasizing that editing preference for UAG over GAC is not derived from differences in binding affinity. Further, all mutants showed a similar binding affinity for UAG and GAC hairpins (Table 2.2). E488Q, T490A and V493T showed similar binding affinity as WT hADAR2, indicating that these mutations do not affect the binding step. However, N597K and N613K showed ~ 2 -fold increase in binding affinity compared to WT

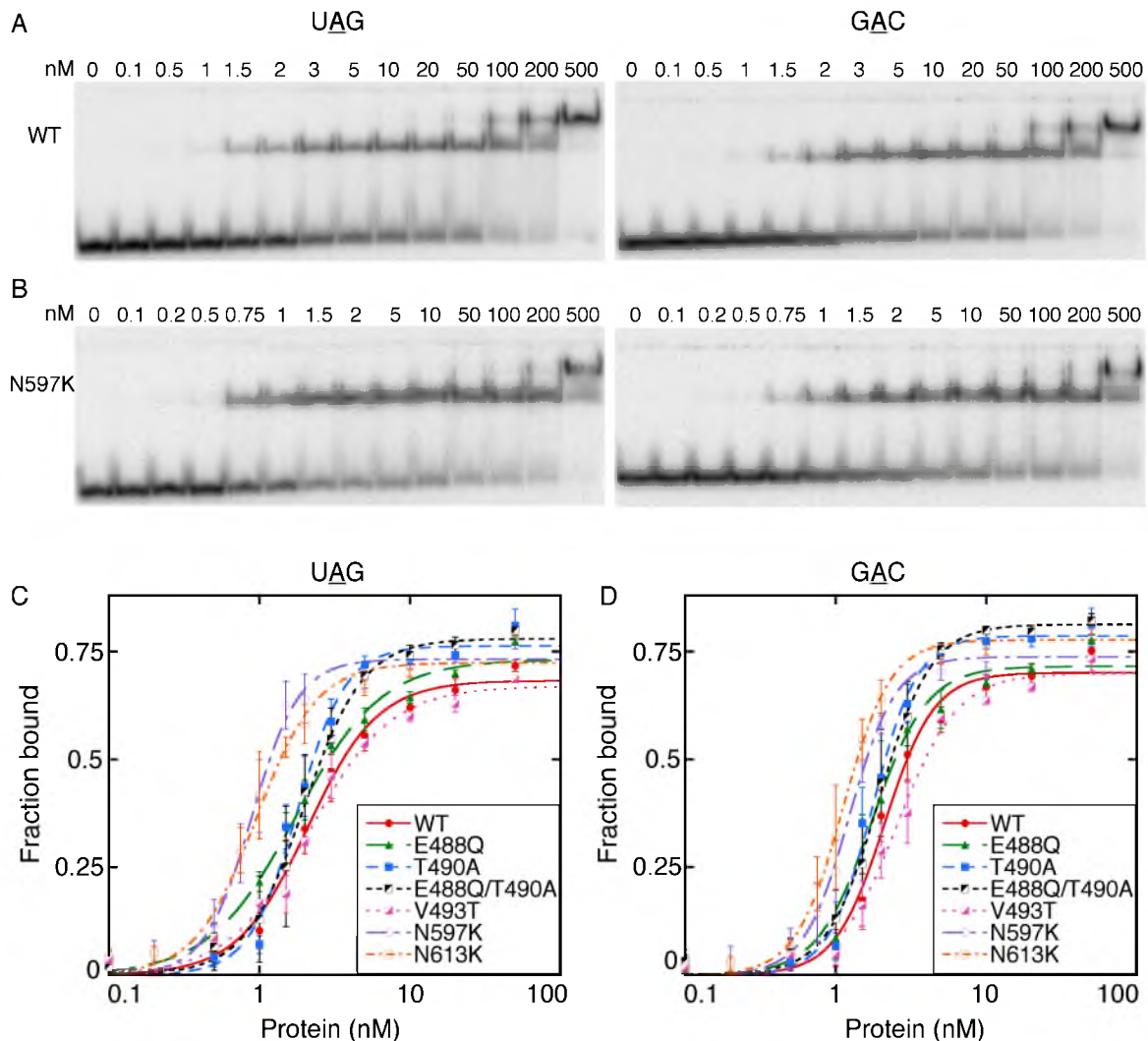


Figure 2.2. Binding affinity is similar to that of wildtype hADAR2 for some mutants, but increased for others. (A and B) PhosphorImages showing representative gel shift assays of WT hADAR2 and the N597K mutant with 20 pM of ^{32}P -5' end-labeled UAG and GAC hairpins (Fig. 2.1B and C). Protein concentrations in nM are indicated at the top of each gel. (C and D) UAG and GAC hairpin binding isotherms for WT hADAR2 and mutant enzymes. Radioactivity corresponding to RNA_{total} and RNA_{free} were quantified to determine the fraction bound [Fraction Bound = $1 - (\text{RNA}_{\text{free}}/\text{RNA}_{\text{Total}})$]. All data points were fit using the Hill formalism. Error bars = standard deviation (s.d., $n \geq 3$).

Table 2.2. Characterization of hADAR2 WT and mutants

hADAR2 proteins	K_d (nM) ^a		k_{deam} (min ⁻¹) ^b		FI (a.u.) ^c		S_{rel} ^d
	UAG	GAC	UAG	GAC	UA _{2AP} G-28	GA _{2AP} C-27	
WT	2.1 ± 0.3	2.1 ± 0.3	0.9 ± 0.1	$(4.4 \pm 0.4) \times 10^{-4}$	4.6 ± 0.4	1.9 ± 0.1	1.00
E488Q	1.8 ± 0.3	1.8 ± 0.3	> ~2.5 ^e	$(2.6 \pm 0.3) \times 10^{-2}$	9.8 ± 0.4	2.6 ± 0.2	0.85
T490A	1.8 ± 0.2	1.8 ± 0.2	$(1.2 \pm 0.1) \times 10^{-2}$	UN ^f	3.9 ± 0.3	2.0 ± 0.1	1.20
E488Q/T490A	2.1 ± 0.4	1.9 ± 0.4	1.0 ± 0.1	$(6.3 \pm 0.4) \times 10^{-3}$	3.9 ± 0.2	2.0 ± 0.1	ND ^f
V493T	2.2 ± 0.1	2.6 ± 0.3	> ~2.5 ^e	$(2.6 \pm 0.3) \times 10^{-3}$	4.2 ± 0.2	1.7 ± 0.1	ND ^{f, g}
N597K	0.9 ± 0.1	1.2 ± 0.1	2.3 ± 0.5	$(2.3 \pm 0.3) \times 10^{-3}$	NR ^f	NR ^f	0.67
N613K	1.0 ± 0.2	1.2 ± 0.1	1.2 ± 0.2	$(1.6 \pm 0.2) \times 10^{-3}$	2.6 ± 0.1	1.6 ± 0.1	0.79

^a K_d , dissociation constant for proteins binding UAG or GAC hairpins.

^b k_{deam} , rate constant for editing underlined adenosine of UAG and GAC hairpins.

^c FI, increase in fluorescence intensity (arbitrary units) on addition of protein to UA_{2AP}G-28 or GA_{2AP}C-27. FI of controls:

ssUA_{2AP}G-28 = 10.2 ± 0.3, duplex UA_{2AP}G-28 = 3.7 ± 0.1, ssGA_{2AP}C-27 = 9.9 ± 0.3, duplex GA_{2AP}C-27 = 3.4 ± 0.2.

^d S_{rel} , Relative nearest neighbor specificity for WT and mutants ($S_{\text{rel}} = S_{\text{protein}}/S_{\text{WT}}$). Nearest neighbor specificity for proteins, for example, WT hADAR2, was determined from the equation: $S_{\text{WT}} = [\sum | \text{Average \% edited for a triplet} - 20 |]_{\text{WT}}$, which is summation of the absolute values obtained by subtracting 20 from the average % edited for each triplet. Total % editing was normalized to 20% (see text and Fig. 2.5).

^e Values measured for E488Q and V493T with UAG were actually 7.3 ± 0.3 min⁻¹ and 3.4 ± 0.3 min⁻¹, respectively. However, in our experience the manual pipetting method we used is inaccurate for values above 2.5 min⁻¹, and a more accurate value must await measurement by rapid quench protocols.

^f UN=Undetectable, NR = Not Relevant and ND = Not Determined.

^g $S_{\text{rel}} = 0.80$ for V493A, a different mutation at this position.

hADAR2, for both UAG and GAC hairpins; this difference was reproducible between different experiments and different protein preparations. Possibly, these mutants were selected in the screen due to increased binding affinity. Consistent with this idea, both residues are on the surface of the protein (21). In the full-length protein used for the gel-shift assays, the dsRBMs likely contributed far more to the observed affinity than the catalytic domain, possibly masking the full impact of the mutations on interactions with the catalytic domain. In fact, a full-length protein containing both mutations, still showed only ~2-fold increase in binding affinity compared to WT hADAR2 (Supplementary Fig. 2.S1A).

However, when we compared truncated proteins, one consisting of the WT catalytic domain (truncWT), and the other consisting of the catalytic domain containing both mutations (truncN597K/N613K), truncN597K/N613K had ~4-fold higher binding affinity than truncWT (Supplementary Fig. 2.S1B and C).

Compared to WT hADAR2, the catalytic rate of E488Q is increased and that of T490A is decreased

WT hADAR2 and variants showed similar binding affinity for the UAG and GAC hairpins, indicating that discrimination between UAG and GAC occurs after the initial binding step. To further understand the basis of preferences, we determined the deamination rate (k_{deam}) of WT hADAR2 and mutants under single turnover conditions. Slow turnover rate and substrate inhibition of ADARs makes steady state measurements challenging (24). We used the same UAG and GAC hairpins for determining deamination rate as was used for binding affinity studies, except that the editing site adenosine was

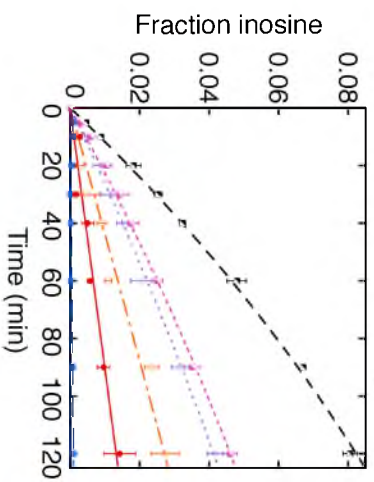
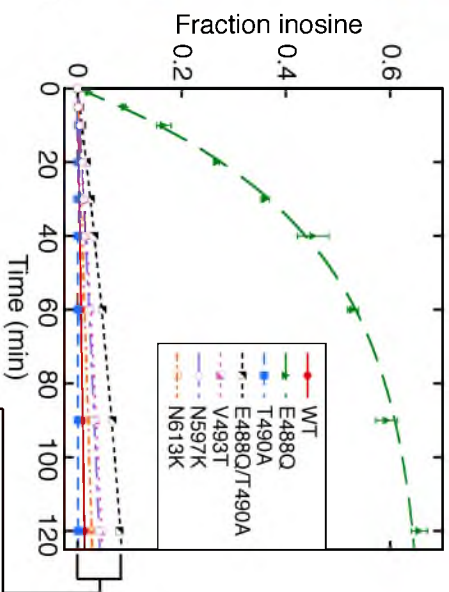
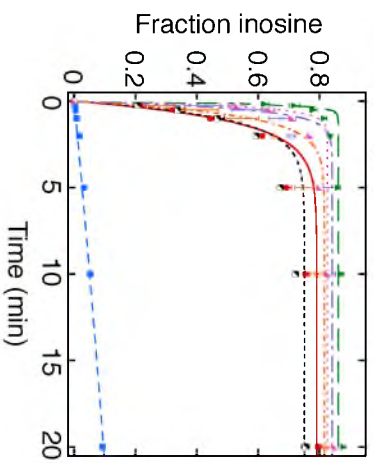
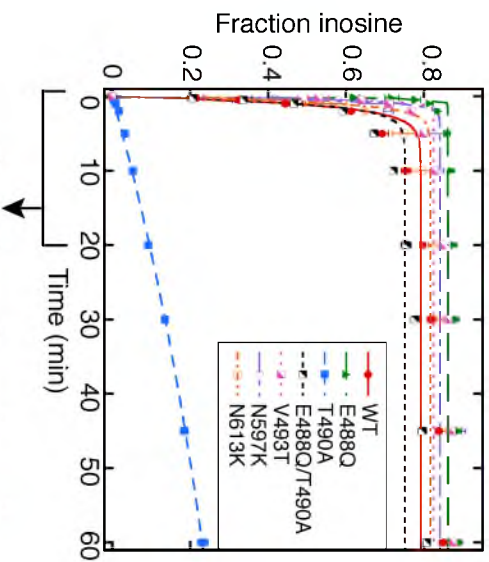
^{32}P labeled at its 5' phosphate using a splint ligation technique (25). Hairpins were incubated with enzyme, and then the RNA was treated with nuclease P1 to produce nucleoside 5'-monophosphates, which were separated by thin layer chromatography (TLC). Deamination rate was determined by monitoring the amount of 5'-AMP converted to 5'-IMP over time.

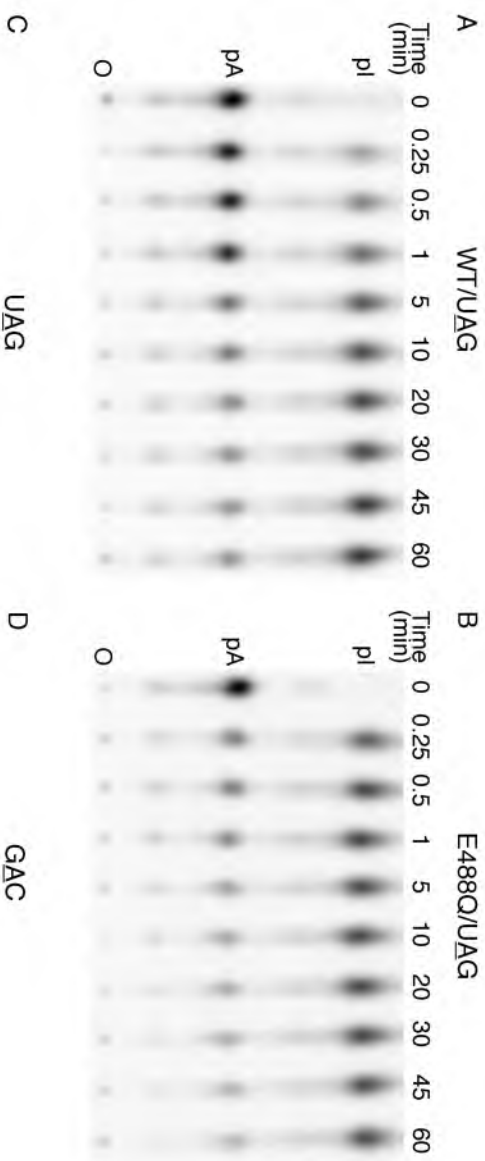
Representative deamination assays for WT and E488Q mutant hADAR2, using the UAG hairpin, are shown (Fig. 2.3A and B), and quantitation of multiple assays for both UAG and GAC hairpins plotted (Fig. 2.3C and D). Compared to WT hADAR2, the E488Q mutant showed an increase in deamination rate for the UAG hairpin (Table 2.2; Fig. 2.3C). V493T and N597K also showed an increase in deamination rate for UAG, albeit to a lesser extent than observed with E488Q (Table 2.2; Fig. 2.3C). Dramatically, the E488Q mutant showed ~60-fold increase in deamination rate for the GAC hairpin compared to WT hADAR2, whereas V493T, N597K and N613K showed only a slight increase in deamination rate for this hairpin (Table 2.2; Fig. 2.3D). These data correlated with *in vivo* data obtained from the screen, which showed that E488Q edited GAC most efficiently. T490A edited the UAG hairpin poorly, and did not edit the GAC hairpin, again correlating with *in vivo* data. These data indicated that T490 is important for wildtype levels of editing, and is involved in a step after the initial binding.

Assays of 2-AP fluorescence suggest certain mutants alter base-flipping

Since the E488Q mutant bound RNA with an affinity similar to that of WT hADAR2, its large increase in deamination rate was likely due to a subsequent step. Like other enzymes that modify bases within a double helix, ADARs are thought to use a base-

Figure 2.3. Deamination rates for some mutant enzymes are similar to that of WT hADAR2, but others differ. (A and B) PhosphorImages showing representative TLC plates used in the deamination rate assay with 250 nM WT hADAR2 or E488Q mutant, and 0.5 nM UAG hairpin with the target adenosine labeled at its 5' phosphate. Time points are indicated at the top of the TLC plate, and on the left, positions of origin (O), 5' AMP (pA) and 5' IMP (pI). Control experiments using less protein or twice the amount of RNA confirmed single-turnover conditions (Supplementary Fig. 2.S2), and also established that WT and mutant hADAR2 were stable for the duration of the experiment (Supplementary Fig. 2.S3). (C and D) Plots showing the fraction inosine produced as a function of time for WT hADAR2 and mutants, with UAG and GAC hairpins. Data points were fit to the equation, $F_t = F_{end} (1 - e^{-kt})$, where F_t is the fraction inosine at time t , F_{end} is the fitted fraction inosine at end point and k is the fitted rate constant. Error bars = s.d. ($n \geq 3$). Inset expands the x-axis for reactions with the UAG hairpin, and y-axis for reactions with the GAC hairpin. While the overall fit to this equation was good, for the UAG hairpin, late time points showed a continued increase in inosine. While this could indicate a double exponential rate, the k_{deam} values obtained on excluding the late time points by fitting the data points up to 30 min were similar to that obtained from 60 min time points. The small increase at later time points was possibly due to slow editing of contaminating ^{32}P 5' end labeled 54 nt RNA, used as starting material for preparing the 60 nt UAG hairpin by splint ligation.

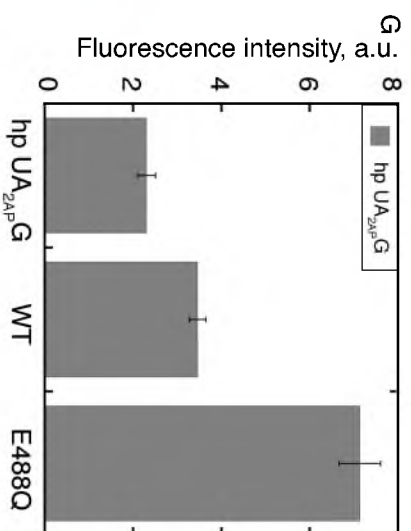
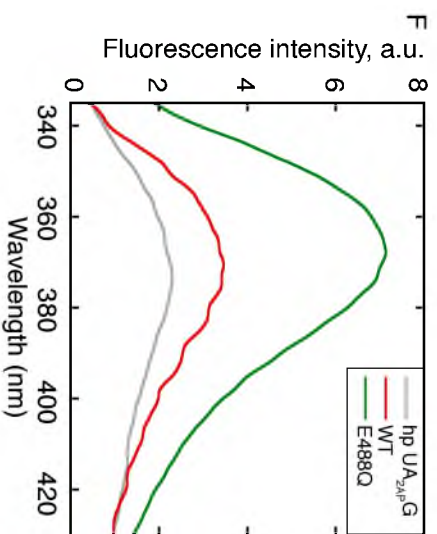
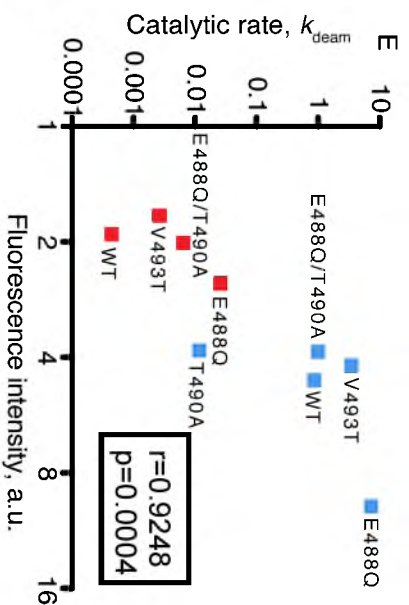
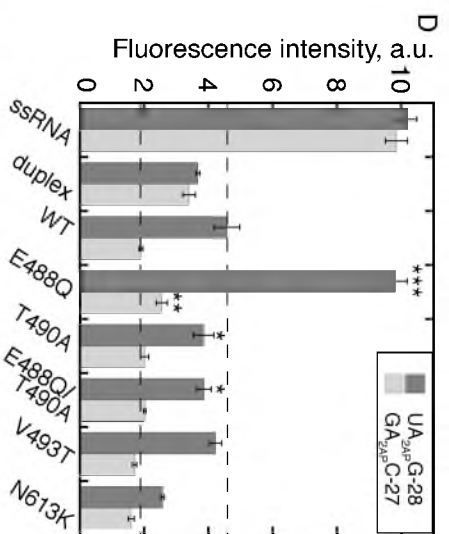




flipping mechanism (24). In another base-flipping enzyme, the cytosine-specific DNA methyl transferase M.HhaI, the position occupied by the target cytosine in the DNA duplex is assumed by a Gln when cytosine flips out (26). This Gln is flanked by Gly residues, proposed to be crucial for positioning the Gln side chain for deep penetration into the helix. E488Q also has flanking Gly residues and is on a loop proximal to the active site. Thus, we investigated the base-flipping ability of mutants on this conserved loop (E488Q, T490A and V493T) by substituting the target adenine with 2-aminopurine (2-AP), a fluorescent adenine analog previously used to probe base-flipping (27), including in studies of hADAR2 (24).

The fluorescence of 2-AP is dependent on its molecular environment. When present in single stranded oligonucleotides, or free in aqueous solution, 2-AP fluoresces. However, when incorporated into a double helix, its fluorescence is quenched due to base-stacking interactions. Two ADAR substrates were synthesized, with 2-AP in the context of favored (UA_{2AP}G-28) or disfavored (GA_{2AP}C-27) neighbors (Fig. 2.4A). These substrates were similar to those used for determining binding affinity and deamination rate, except the editing site adenine was replaced with 2-AP, and intermolecular duplexes were used instead of a hairpin, to enable control experiments with single stranded RNA. As expected, in control experiments, single stranded RNA with 2-AP showed a dramatic increase in fluorescence intensity (FI) compared to duplex RNA with 2-AP (grey dashed and solid lines, Fig. 2.4B and C). However, the 2-AP fluorescence in the duplex was not completely quenched, possibly because the fluorescence of 2-AP in a mismatch is less effectively quenched than 2-AP in a base pair (27). In the absence of protein, UA_{2AP}G-28 and GA_{2AP}C-27 duplexes showed similar FI,

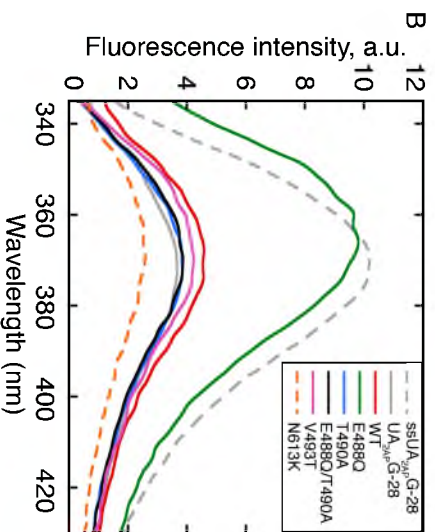
Figure 2.4. 2-AP fluorescence assays suggest certain mutants alter base-flipping. (A) Sequences of constructs used for base-flipping studies, with target adenosines substituted by 2-AP (red). UA_{2AP}G-28 and GA_{2AP}C-27 are intermolecular duplexes made from complementary strands of 28 and 27 nts respectively. (B and C) Plots showing FI in arbitrary units (a.u.) as a function of wavelength, for samples containing only RNA (0.6 μ M), or both protein (2.4 μ M) and RNA (0.6 μ M) for UA_{2AP}G-28 and GA_{2AP}C-27. Excitation was at 320 nm to minimize background fluorescence from excitation of protein residues, and emission was scanned from 335-430 nm. Each spectrum is the average of multiple analyses ($n \geq 3$), and mean FI at emission maximum and s.d. (error bars) is plotted in (D). (D) Dotted lines indicate observed FI of WT hADAR2 with UA_{2AP}G-28 and GA_{2AP}C-27 for reference. Asterisks indicate p-values for the mutants compared to WT hADAR2 (* p-value = 0.02, ** p-value = 0.007, *** p-value = 0.00002). (E) Plot of Pearson product-moment correlation coefficient measuring correlation between catalytic rate and FI increase observed with UAG or GAC substrates for WT hADAR2 and mutants. UAG substrate data, blue; GAC substrate data, red. Correlation coefficient (r) and p-value (p) are indicated. The correlation coefficients for UAG substrate data only or GAC substrate data only were also greater than 0.91. (F and G) Plot and bar graph showing FI with UA_{2AP}G hairpin, analyzed similarly to UA_{2AP}G - 28 duplex.



A

UA_{2AP}-G-28

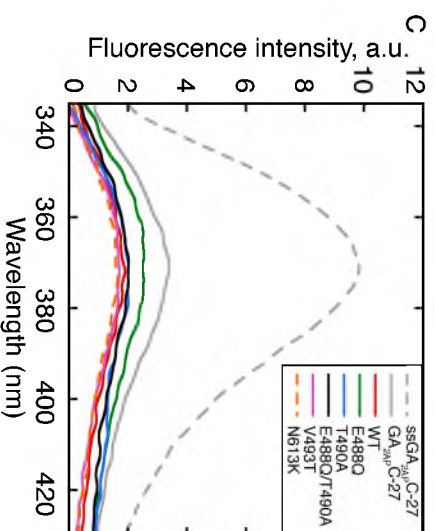
5'-CCGUUU^{2-AP}GGUGGUA^GUAUACCA
3'-GGCAGACCAUCCACC^CUAU^GAUAUGG



C

GA_{2AP}-C-27

5'-CCGUUU^{2-AP}CUGGGUA^AUAUACC
3'-GGCAGACCAUCCACC^CUAU^GAUAUGG



as did single stranded UA_{2AP}G-28 and GA_{2AP}C-27 (Fig. 2.4D).

All fluorescence measurements were made at saturating protein concentrations. When WT hADAR2 was added to the UA_{2AP}G-28 duplex, in agreement with previous studies (24), FI increased compared to that observed with duplex UA_{2AP}G-28 alone (p-value = 0.02; Fig. 2.4B, compare solid red and grey lines). Most notably, E488Q showed a dramatic increase in FI compared to WT hADAR2 (p-value = 0.00002; Fig. 2.4B, solid green versus red lines), suggesting that a Gln at residue 488 enhances base-flipping. Our binding studies indicated that T490A has a wildtype affinity for dsRNA (Table 2.2), but the FI observed when this protein was added to UA_{2AP}G-28 was indistinguishable from that of the duplex alone (Fig. 2.4B, solid blue and grey lines). This suggests that T490 might be required for base-flipping or in a step upstream of base-flipping. V493T did not show a statistically significant difference in FI compared to WT hADAR2 (Fig. 2.4B, solid pink versus red lines).

Surprisingly, when added to duplex GA_{2AP}C-27 (Fig. 2.4C, grey solid line), WT hADAR2 and all mutants showed a decrease in FI compared to duplex GA_{2AP}C-27 alone. Yet, the FI observed with E488Q was higher than that observed with WT hADAR2 (green versus red lines, p-value = 0.007), suggesting that E488Q also enhances base-flipping of adenosine within GAC. These data suggest that the net FI observed in our steady-state fluorescence measurements reflects both quenching due to protein binding at the mismatched 2-AP, as well as base-flipping. According to this hypothesis, for WT hADAR2, E488Q and V493T, base-flipping with the UA_{2AP}G-28 duplex is more robust than that occurring with GA_{2AP}C-27 duplex, leading to an increase in FI that

counterbalances the quenching due to protein binding, producing a net increase in FI (Fig. 2.4B). Accounting for both binding and base-flipping in the net FI, might also explain the relatively small increase in FI when WT hADAR2 was added to the UA_{2AP}G-28 duplex, and results with N613K, a mutant with ~2-fold higher binding affinity and similar catalytic rate as WT hADAR2. Compared to WT hADAR2, N613K showed ~2-fold decrease in FI with UA_{2AP}G-28 duplex (compare red and dashed orange lines, Fig. 2.4B; Table 2.2). Finally, given that T490A exhibits low levels of editing with UAG substrates, it likely is capable of base-flipping, albeit inefficiently. We presume that effects of this base-flipping on FI are counterbalanced by quenching, so that FI with T490A and the UA_{2AP}G-28 duplex is indistinguishable from that of the duplex alone.

While our FI measurements likely reflect contributions from both binding and base-flipping, for proteins with similar affinity, an increase in FI should correlate with increased base-flipping. Consistent with this, we observed a positive correlation between increase in FI and catalytic rate for WT and hADAR2 variants with similar affinity (Pearson's product moment correlation coefficient (r) = 0.9248, p -value = 0.0004; Fig. 2.4E). To confirm that fluorescence experiments performed with duplexes could be compared to experiments using hairpins, we incorporated a 2-AP at the editing site of the UAG hairpin used in binding and deamination assays. When the UA_{2AP}G hairpin was mixed with WT hADAR2 and E488Q, we observed an increase in FI comparable to that observed with the UA_{2AP}G-28 duplex (Fig. 2.4F and G).

Characterization of the double-mutant E488Q/T490A

Compared to WT hADAR2, the E488Q mutant showed a dramatic increase in FI when added to UA_{2AP}G-28 duplex, suggesting that Gln at residue 488 affects base-flipping. In contrast, when added to this duplex, the T490A mutant exhibited a FI that was slightly less than that observed with WT hADAR2 (difference significant at p-value=0.02), suggesting that T490 is required for efficient base-flipping or in a step prior to base-flipping. We hypothesized that if T490 was essential for efficient base-flipping or in a step prior to base-flipping, then the double-mutant, E488Q/T490A, would not exhibit the dramatic increase in FI observed with the E488Q mutant.

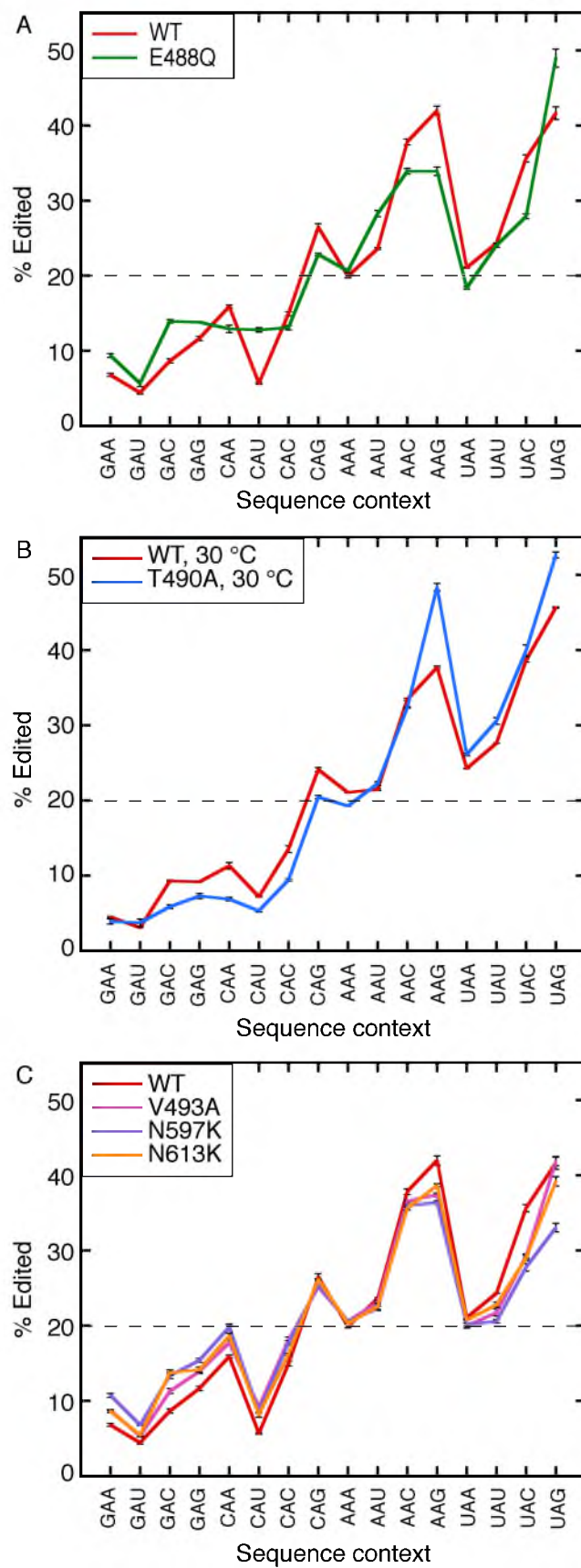
We first confirmed that binding affinity of E488Q/T490A was similar to WT hADAR2 for both UAG and GAC hairpins (Fig. 2.2C and D; Table 2.2). While the catalytic rate of the T490A mutant was extremely low for both hairpins, in the context of the E488Q/T490A double-mutant, deamination rate increased for both hairpins, but was far less than the high deamination rate exhibited by the E488Q single mutant (Fig. 2.3C and D; Table 2.2). Further, when the E488Q/T490A mutant was mixed with UA_{2AP}G-28 or GA_{2AP}C-27 duplex, the increase in FI was similar to that of T490A, indicating that T490 is essential for the increase in FI observed with E488Q (Fig. 2.4B, C and D; Table 2.2). The E488Q mutation could not enhance base-flipping of the T490A mutation, but the increased deamination rate of the double-mutant compared to the T490A mutant suggested E488Q has an additional role in catalysis.

hADAR2 specificity is affected by the E488Q and T490A mutations

With WT hADAR2, a protein induced increase in FI was observed with UA_{2AP}G-28 duplex compared to GA_{2AP}C-27 duplex, suggesting that hADAR2's preference for UAG over GAC is derived from differences in base-flipping. Are there other examples where the base-flipping efficiency of the enzyme affects specificity? In M.EcoRI, an N6-adenine DNA methyl transferase that utilizes a bending, base-flipping and intercalation mechanism, a bending deficient mutant decreases base-flipping and increases specificity (28, 29). For noncognate substrates, M.EcoRI specificity arises from partitioning the enzyme/DNA intermediate into the unbent form (29, 30). We investigated whether E488Q and T490A, mutants that showed differences in 2-AP FI compared to WT hADAR2, also showed differences in substrate specificity. We determined nearest neighbor preferences for all possible sixteen triplet contexts in which the target adenosine can occur using a long synthetic, perfectly base paired dsRNA (19). Each protein was incubated with non-radiolabeled 418 bp dsRNA to achieve ~20% overall editing; this ensured that well edited sites were not saturated to 100% editing, which could result in loss of information, and also that the majority of editing sites with a 5' A were not 5' I (19), which could skew preference determinations.

If hADAR2 lacked preferences, the graphs in Fig. 2.5 would each show a horizontal line at 20% editing (see dotted line). However, as illustrated for the WT protein, hADAR2, like all ADARs, exhibits preferences, and the line graph is plotted with the most preferred triplets to the right of the graph. Consistent with the idea that increased base-flipping correlates with a decrease in specificity, for E488Q, most points on the line graph moved closer to the dotted line representing 20% editing (Fig. 2.5A).

Figure 2.5. hADAR2 mutations affect specificity. (A) Plot showing average % editing of adenosine in each of the 16 possible triplet contexts, determined from analysis of editing in a 418 bp dsRNA, for E488Q compared to WT hADAR2. 196 adenosines were used to calculate average % editing of adenosine in 16 triplet contexts. Triplets are ordered in terms of 5' nearest neighbor, with 5' G followed by C, A and U. Mean % editing across 196 adenosines was 19.9% for WT hADAR2 and 21.3% for E488Q, and was normalized to 20% as indicated by the dotted line. Error bars = s.d. ($n \geq 3$). (B) Similar plot as in (A), comparing T490A to WT hADAR2. Sequencing for a portion of the antisense strand was not clean for the T490A reactions, so only 153 adenosines were used to calculate average % editing of adenosine in 16 triplet contexts for both T490A and WT hADAR2. Mean % editing across 153 adenosines was 18.7% for WT hADAR2 and 19.2% for T490A and was normalized to 20% as in (A). Error bars = s.d. ($n \geq 2$). (C) As in (A), comparing V493A, N597K and N613K to WT hADAR2. Mean % editing across 196 adenosines was 19.6% for V493A, 19.0% for N597K and 20.6% for N613K and was normalized to 20% as in (A). Error bars = s.d. ($n \geq 3$).



Overall, the E488Q mutant showed more editing for the least preferred triplets (left half of the graph) and less editing for the most preferred triplets (right half of the graph), although some triplets (CAA, CAC, AAU, UAG) did not follow this pattern. To facilitate comparison between enzyme variants, we defined a relative nearest neighbor specificity, S_{rel} , and compared to the WT enzyme, the E488Q enzyme exhibited a $S_{rel} = 0.85$ (Table 2.2). As expected, the E488Q mutant showed a slight to moderate increase in editing of all four triplets that had a 5' G. However, compared to WT hADAR2, only a 1.6-fold increase in editing of GAC was observed with E488Q in the 418bp dsRNA, in contrast to ~60-fold increase in deamination rate observed in the GAC hairpin. Possibly this disparity indicates that the mismatched adenosine of the GAC hairpin is more amenable to base-flipping than adenosines within the completely base-paired 418 bp dsRNA.

The T490A mutant, on the other hand, showed an overall increase in specificity, with $S_{rel} = 1.20$ (Table 2.2); triplets poorly edited by WT hADAR2 were edited to a lower percentage, and triplets well edited by WT hADAR2 were edited to a greater percentage (Fig. 2.5B). Mutants V493A, N597K and N613K showed a decrease in specificity (Table 2.2); triplets poorly edited by WT hADAR2 were edited more, and triplets well edited by WT hADAR2 were edited less (Fig. 2.5C). The decreased specificity observed with N597K and N613K could be due to their higher binding affinity compared to WT hADAR2.

Discussion

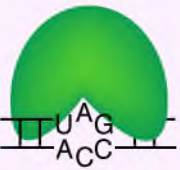
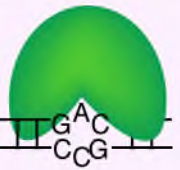
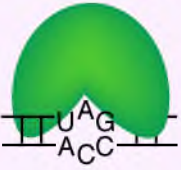
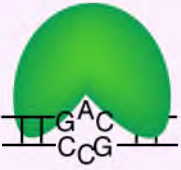
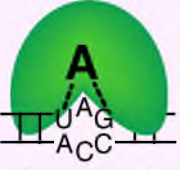
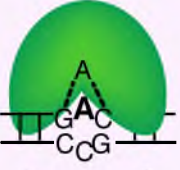
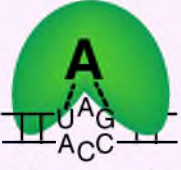
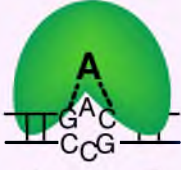
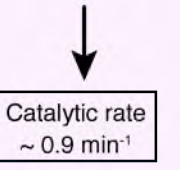
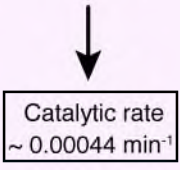
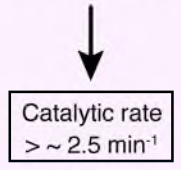
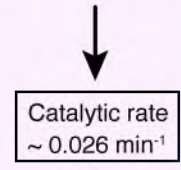
ADARs exhibit nearest neighbor preferences in choosing adenosines for deamination. Preferences derive mainly from the catalytic domain (19, 20), but the amino

acids that mediate preferences, and the mechanism involved, are unclear. We investigated these issues by performing a screen for mutations within the catalytic domain that allowed editing of an adenosine within a disfavored nearest neighbor context, GAC. We identified seven mutations, most of which mapped onto two distinct regions on the surface of the predicted RNA binding site (21). We characterized effects of these mutations by performing *in vitro* assays on mutant hADAR2 enzymes, to determine binding affinity, catalytic rate, and nearest neighbor preferences. Using a hADAR2 substrate with the fluorescent analog 2-AP incorporated at the editing site, we also compared changes in fluorescence that occurred upon addition of WT or mutant enzymes. These data support the idea that ADARs use a base-flipping mechanism to access the target adenosine, and unexpectedly, suggest that preferences derive mainly from nearest neighbor effects on base-flipping, rather than direct recognition of neighboring bases. Our studies provide the first information in regard to the region of hADAR2 important for preferences, pointing to a conserved loop on the surface of the protein and close to the active site.

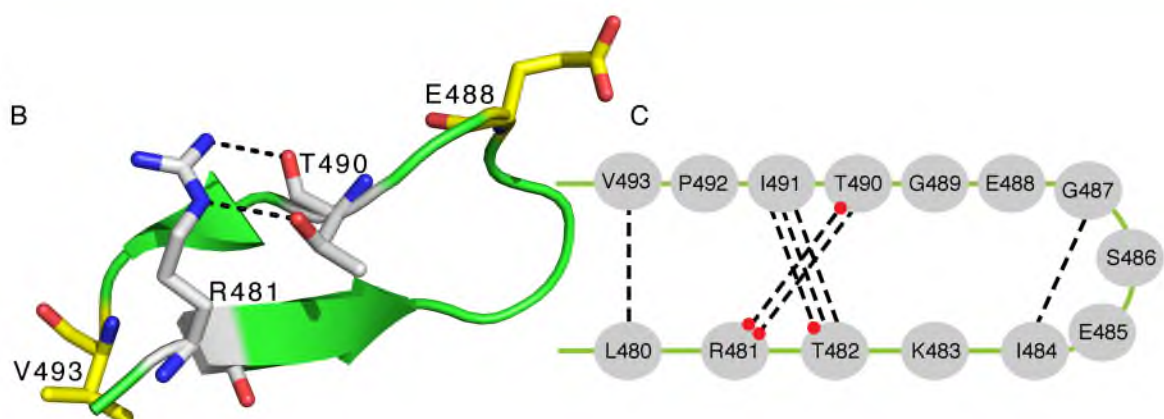
Fig. 2.6 presents a model that incorporates the results of our studies. The model compares WT hADAR2 with the E488Q mutant, but is consistent with properties of all characterized mutations. As illustrated, we observed that WT hADAR2 and the E488Q mutant bound the favored UAG and disfavored GAC substrates with identical affinities; this was a consistent trend for all mutant enzymes. Even for the two mutants found to have a higher affinity compared to the WT enzyme (N597K, N613K), no differences were observed between UAG and GAC substrates. Thus, our data indicate that preferences do not derive from differential binding. Our model posits that a base-flipping

Figure 2.6. Model describing preference for adenosine in context of UAG compared to GAC, and the role of residues in the conserved loop. (A) Three steps, binding, base-flipping and editing are shown for WT hADAR2 and E488Q with UAG and GAC substrates. Binding affinities, increase in FI upon mixing enzymes with 2-AP substituted UAG or GAC substrates, and catalytic rates are indicated. Binding affinities of WT hADAR2 and E488Q are similar for UAG and GAC substrates, indicating that discrimination is not derived from differences in binding affinity. In the second step, base-flipping is more efficient for adenosine within UAG than within GAC (indicated by font size), correlating with increased catalytic rate (third step), and suggesting preferences derive from differences in base-flipping. Additionally, E488Q facilitates base-flipping, leading to a further increase in catalytic rate for E488Q compared to WT hADAR2. In our model, T490 is essential for stability of the active conformation of the conserved loop, and a residue on this loop is important for base-flipping. (B) A close view of the conserved loop that includes two β strands and comprises 14 residues. Hydrogen bonds between the R481 side chain and T490 backbone carbonyl oxygen and side chain hydroxyl are shown. Residues E488 and V493 on this loop are also indicated (yellow sticks). (C) Cartoon showing all 7 hydrogen bonds within the 14 residues as determined in the crystal structure (21). Red dots indicate a hydrogen bond from a side chain, with others involving a backbone carbonyl oxygen or amine hydrogen.

A

	WT hADAR2		E488Q	
Binding Binding similar				
	K_d 2.1 ± 0.3 nM	K_d 2.1 ± 0.3 nM	K_d 1.8 ± 0.3 nM	K_d 1.8 ± 0.3 nM
Base-flipping Adenosine within UAG flips more than that in GAC E488Q aids in base-flipping T490 essential for efficient base-flipping				
	FI 4.6 ± 0.4	FI 1.9 ± 0.1	FI 9.8 ± 0.4	FI 2.6 ± 0.2
Deamination				
	Catalytic rate ~ 0.9 min ⁻¹	Catalytic rate ~ 0.00044 min ⁻¹	Catalytic rate $> \sim 2.5$ min ⁻¹	Catalytic rate ~ 0.026 min ⁻¹

B



step follows binding, and here our experiments with 2-AP suggest a clear difference between UAG and GAC substrates. For WT hADAR2 and all studied mutants, a greater increase in fluorescence was observed when the protein was mixed with UAG substrates compared to GAC substrates, and in all cases this correlated with a higher deamination rate (Fig. 2.4D and E). The change in fluorescence was most dramatic with the E488Q mutant, suggesting this mutant has enhanced base-flipping properties. Our data suggest preferences are based on effects of nearest neighbors on the base-flipping step of the ADAR reaction.

We also characterized residue T490, since it was on the conserved loop, proximal to other residues identified in the screen (Fig. 2.1F and G; Fig. 2.6B). The mutant T490A exhibited a greatly reduced catalytic rate (Table 2.2), and minimal base-flipping that was not enhanced in the E488Q/T490A double-mutant. These data suggest that T490 is required for efficient base-flipping, and our favored model is that it is important for maintaining a conformation of the conserved loop that promotes base-flipping. Indeed, the hADAR2 crystal structure shows the side chain hydroxyl and backbone carbonyl of T490 hydrogen bonding with the side chain of R481 (21) (Fig. 2.6B and C). Correspondingly, R481A did not show editing of adenosine within UAG or GAC hairpins in the *in vivo* α -galactosidase reporter assay (Supplementary Fig. 2.S5). Further, in the mutational analysis of residue T490 (Table 2.1), only Ser and Cys could replace Thr for editing UAG, suggesting that the predicted hydrogen bond (21) from the side chain of T490 to R481 is important (Fig. 2.6B).

The double-mutant E488Q/T490A showed a similar minimal level of base-flipping as T490A, but a higher deamination rate than T490A. This suggests that a Gln at 488 has

effects in addition to enhancing base-flipping. Similar to another base-flipping enzyme, M.EcoRI, where enhanced base-flipping compromises specificity (28), the enhanced base-flipping of the E488Q mutant correlated with a decreased specificity for most triplets (Fig. 2.5A); however, some triplets did not conform, again raising the possibility that the E488Q mutation has other effects. Compared to WT hADAR2, only Gln or Asn at 488 enhanced GAC editing (Table 2.1), emphasizing the importance of an amide side chain for GAC editing. Possibly a hydrogen bond from the amide side chain to the RNA promotes this specificity. With E488Q, a small increase in base-flipping of adenosine within GAC (FI ~2.6) resulted in ~60 fold increase in editing. This might also indicate a Gln at 488 has effects beyond base-flipping, but it is also possible that for GAC, base flipping is rate limiting, and a small increase in base-flipping results in a large increase in catalytic rate.

2-AP fluorescence is affected by changes in the immediate environment, such as alterations in base-pairing or protein interactions (31). The analog is frequently used to report on base-flipping (27), including with hADAR2 (24), and in limited cases the observed fluorescence changes have been correlated with a crystal structure of 2-AP in a flipped out position (32). While we cannot be certain our assays with 2-AP are measuring base-flipping, this seems the most likely explanation. The C6 position of adenine that undergoes nucleophilic attack during the deamination reaction lies deep in the major groove of dsRNA, and as previously proposed (15, 24), it is difficult to imagine anything but base-flipping that would allow ADAR access to the C6 atom. Further, the crystal structure of the hADAR2 catalytic domain shows a large basic patch on the surface, which likely facilitates binding to dsRNA, with a deep pocket that contains the active site

zinc ion (21), a perfect arrangement for interacting with a helix with a flipped out adenosine.

Proven base-flipping enzymes utilize various mechanisms to gain access to a base. These include directly pushing it out of the helix with an amino acid (33), the serine mediated pinch-pull-push mechanism (34), or the helix bending, base-flipping and intercalation mechanism (30). An alternative passive mechanism has also been proposed, where the protein simply traps a transiently flipped out base (35). Our studies suggest that hADAR2 base-flips an adenosine within a UAG context more efficiently than within a GAC context, but further studies will be necessary to determine mechanistic details. If ADARs use the passive mechanism, it implies that an adenosine with a 5' U intrinsically flips more readily than an adenosine with a 5' G. Indeed, NMR experiments and theoretical calculations indicate base pair opening probability is affected by nearest neighbors (36, 37), and data from such analyses are consistent with ADAR preferences. For example, the opening probability of an A•T base pair with a 5' G is less than that with a 5' C or 5' T, possibly due to the stronger stacking interaction of adenosine with a 5' G (37). Alternatively, base-flipping may be protein induced, but the extent of flipping might be influenced by the sequence context. Molecular dynamics simulations suggest that, of the two A•C mismatches in the R/G hairpin, adenosine of the target A•C pair is more prone to base-flipping than that in the second A•C pair, which is not edited (38). Interestingly, the target adenosine has a 5' A, while the adenosine in the second A•C pair has a 5' G.

The locus for dyschromatosis symmetrica hereditaria (DSH), a pigmentary genodermatosis, maps to the *hADAR1* gene (9). Out of more than 100 mutations reported

in the *hADAR1* gene of DSH patients, at least 47 are missense mutations in the catalytic domain (9, 39). DSH is not usually associated with other diseases, but in the two reported cases where DSH is accompanied by dystonia, brain calcification and mental deterioration, the same mutation, G1007R, was identified (40, 41). The equivalent residue in hADAR2 is G487, one of the two Gly residues flanking E488 on the conserved loop. It will be informative to study this mutant further to determine if disease symptoms result from altered hADAR1 specificity.

Materials and methods

Construction of hADAR2 plasmids for the screen

The hADAR2 construct includes an N-terminal 12-histidine tag followed by a TEV protease recognition site in a yeast expression plasmid YEpTOP2PGAL1 (42). We modified this construct to include a restriction site before the catalytic domain (YEpTOP2PGAL1-RSinADAR2; Supplementary Methods). Both constructs were cloned into the YCp (yeast centromere plasmid) vector with a *GAL* promoter and *URA3* marker (YCp-ADAR2 and YCp-RSinADAR; Supplementary Methods).

Construction of hairpin-reporter strains

The plasmid encoding UAG hairpin upstream of the α -galactosidase reporter, pR/G α Gal, has been described (22). We constructed the plasmid encoding the GAC hairpin-reporter by performing sewing PCR, and incorporated the favored six nucleotide loop (22) (Supplementary Methods). Both hairpin-reporter constructs were cloned separately into the YCp vector with the *ADHI* promoter and *TRP1* marker, then cloned

into the corresponding YIp (Yeast Integrating Plasmid) vector (Supplementary Methods). YIp vectors containing hairpin-reporters were linearized and integrated into the W303 α chromosome by lithium acetate transformation (Supplementary Methods).

Screen

YCp-RSinADAR2 was restriction digested to produce a gapped vector lacking the catalytic domain (Supplementary Methods). hADAR2 catalytic domain was PCR amplified from YCp-ADAR2 (Supplementary Methods), and used as a template for random mutagenesis by error prone PCR using the GeneMorph II Random Mutagenesis Kit (Stratagene). Amount of DNA template and number of PCR cycles were optimized to get 0-4 mutations per kilobase. 100 ng of the gapped vector and 300 ng of random mutagenized PCR product were simultaneously transformed with lithium acetate into the GAC hairpin-reporter yeast strain, plated on complete minimal (CM)-URA agar plates, and incubated at 30 °C for 2 days. Plates were replica plated onto agar plates containing CM-URA, 3% galactose, 2% raffinose and 0.06 mg/ml X- α -Gal, and incubated at 20 °C for 4-5 weeks. Plasmids were rescued from selected colonies, retransformed into fresh GAC hairpin-reporter yeast strains to confirm hADAR2 dependence, and sequenced to identify mutations.

Additional mutational analysis for selected residues used the same hADAR2 catalytic domain template as used for random mutagenesis. We performed sewing PCR using two outside primers (CDRanMutP1, CDRanMutP3) and two inside primers specific for each mutant (Supplementary Table 2.S1). Inside primers were degenerate, with the three nucleotides coding for the relevant amino acid randomized. Mutagenized PCR

product and gapped vector were transformed into hairpin-reporter yeast strains, as for the screen, and analyzed similarly.

Protein purification

Mutants in the YCp vector were cloned back into the yeast expression plasmid, YE_pTOP2PGAL1 (Supplementary Methods). WT hADAR2 and all mutants (in YE_pTOP2PGAL1) were purified as described (42) to greater than 97% purity, as determined by Coomassie staining. Identity of purified proteins was confirmed by mass spectrometry.

Preparation of RNA for *in vitro* studies

RNA for *in vitro* studies was chemically synthesized and gel-purified after denaturing PAGE (Supplementary Methods). 5' end labeled hairpins for binding experiments were prepared as described (25) (Supplementary Methods). Internally radiolabeled hairpins for rate determination were prepared by splint ligation as described (24) (Supplementary Methods). Duplex or hairpins with the target adenine replaced with 2-AP were prepared as described (24) with modifications. Purified top and bottom strands were dissolved in hybridization buffer (10 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 50 mM KCl), heated to 95 °C for 4 mins, slow cooled to room temperature for 2hrs, ethanol precipitated and gel-purified after 15% native PAGE. Internally radiolabeled, and nonradiolabeled, 418 bp RNA were synthesized as described (19) (Supplementary Methods). This dsRNA has a 21 nt overhang at each 5' terminus. On the 3' termini, sense strands had a 12 nt overhang and antisense strands had a 13 nt overhang.

Gel mobility shift assay

Gel shift assays were as described (23) with modifications. Assays were performed with 20 pM RNA and varying protein concentrations, incubated at 4 °C for 20 mins in buffer containing 14 mM Tris-HCl (pH 8), 130 mM KCl, 10% glycerol, 1 mM DTT, 100 µg/ml BSA, 0.2 mM β-mercaptoethanol, 20 mM NaCl. Reaction were stopped by loading 10 µl of the reaction directly onto a 6 % (37.5:1 acrylamide/ bis-acrylamide) native gel running at 150V, at 4 °C in 0.5x TBE; gels were electrophoresed for 2 more hrs, dried or frozen, and autoradiographed. For truncated proteins, gel shifts were performed as for full-length proteins except 35 mM salt and 29:1 acrylamide/ bis-acrylamide was used.

Deamination assay

Deamination assays were performed under single turnover conditions as described (25), using 0.5 nM RNA and 250 nM protein. Initial experiments with WT hADAR2 and UAG substrate at 30 °C resulted in ~60% editing at 20 sec, making it challenging to discern differences. Thus, for UAG substrate, all incubations were at 20 °C to slow the reaction. For the GAC substrate, incubations were at 30 °C.

2-AP fluorescence assay

2-AP fluorescence experiments were performed on a Perkin Elmer LS 50 Luminescence Spectrometer at room temperature. Excitation was at 320 nm, emission was scanned from 335-430 nm, and a 10 nm slit width was used for excitation and emission. Fluorescence was measured using a ultra micro cuvette from Hellma (30 µl) with 2.4 µM protein and 0.6 µM RNA in buffer containing 16 mM Tris-HCl (pH 7.5), 8

mM Tris (pH 8), 20 mM KCl, 40 mM NaCl, 8% glycerol, 1 mM DTT (Roche), 0.01% Nonidet P-40 and 0.4 mM β -mercaptoethanol (Sigma). Control experiments used 0.6 μ M duplex or single stranded RNA. Corrections were made to emission spectra to account for fluorescence from protein and buffer; for RNA only samples, spectrum of buffer was subtracted, and for protein-RNA samples, spectrum of protein in buffer was subtracted. To ascertain that saturating protein concentrations were used, we performed control experiments using 3.0 and 1.8 μ M protein instead of 2.4 μ M (Supplementary Fig. 2.S4).

Preference assay

An initial time course was performed as described (19) with internally radiolabeled 418 bp dsRNA, to determine time required for ~20% editing. Preference assays were as described (19) by incubating 0.25 nM non-radiolabeled 418 bp dsRNA and 250 nM protein, at 20 °C, for time required to achieve ~20% editing (Supplementary Methods). For T490A, incubations were at 30 °C, since 20% editing was not achieved at 20 °C by 1 hr. For comparison, WT hADAR2 preferences were determined at 30 °C, and found to be identical to those at 20 °C.

Acknowledgments

We thank S. Pokharel and P. Beal for the plasmid pR/G α Gal, and D. Stillman for the W303 α yeast strain, YCp and YIp vectors, and essential advice, guidance and patience. We thank D. Winge for access to the Perkin Elmer LS 50 Luminescence Spectrometer, B. Schackmann and M. Hanson at the UU DNA/Peptide core facility for RNA and primer synthesis, and C. Nelson and K. Parsawar at the UU Mass Spectrometry

and Proteomics core facility for mass spectrometry of mutant ADARs. We thank J.M. Eggington, H.L. Schubert, and the Bass lab for helpful discussions. We also thank A.A. Krauchuk and P.J. Aruscavage for technical assistance. Shared resources were supported by P30CA042014 from the National Cancer Institute. This work was supported by funds from the National Institute of General Medical Sciences to B.L.B. (R01GM044073).

References

1. Bass BL (2002) RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem* 71:817-846.
2. Nishikura K (2010) Functions and regulation of RNA editing by ADAR deaminases. *Annu Rev Biochem* 79:321-349.
3. Hundley HA & Bass BL (2010) ADAR editing in double-stranded UTRs and other noncoding RNA sequences. *Trends Biochem Sci* 35(7):377-383.
4. Wulff BE & Nishikura K (2010) Substitutional A-to-I RNA editing. *Wiley Interdiscip Rev RNA* 1(1):90-101.
5. Warf MB, Shepard BA, Johnson WE, & Bass BL (2012) Effects of ADARs on small RNA processing pathways in *C. elegans*. *Genome Res*:In press.
6. Samuel CE (2011) Adenosine deaminases acting on RNA (ADARs) are both antiviral and proviral. *Virology* 411(2):180-193.
7. Chen CX, *et al.* (2000) A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains. *RNA* 6(5):755-767.
8. Melcher T, *et al.* (1996) RED2, a brain-specific member of the RNA-specific adenosine deaminase family. *J Biol Chem* 271(50):31795-31798.
9. Maas S, Kawahara Y, Tamburro KM, & Nishikura K (2006) A-to-I RNA editing and human disease. *RNA Biol* 3(1):1-9.
10. Kwak S & Kawahara Y (2005) Deficient RNA editing of GluR2 and neuronal death in amyotrophic lateral sclerosis. *J Mol Med (Berl)* 83(2):110-120.

11. Vollmar W, *et al.* (2004) RNA editing (R/G site) and flip-flop splicing of the AMPA receptor subunit GluR2 in nervous tissue of epilepsy patients. *Neurobiol Dis* 15(2):371-379.
12. Zhang XJ, *et al.* (2003) Identification of a locus for dyschromatosis symmetrica hereditaria at chromosome 1q11-1q21. *J Invest Dermatol* 120(5):776-780.
13. Miyamura Y, *et al.* (2003) Mutations of the RNA-specific adenosine deaminase gene (DSRAD) are involved in dyschromatosis symmetrica hereditaria. *Am J Hum Genet* 73(3):693-699.
14. Bass BL (1997) RNA editing and hypermutation by adenosine deamination. *Trends Biochem Sci* 22(5):157-162.
15. Polson AG & Bass BL (1994) Preferential selection of adenosines for modification by double-stranded RNA adenosine deaminase. *EMBO J* 13(23):5701-5711.
16. Nishikura K, *et al.* (1991) Substrate specificity of the dsRNA unwinding/modifying activity. *EMBO J* 10(11):3523-3532.
17. Lehmann KA & Bass BL (1999) The importance of internal loops within RNA substrates of ADAR1. *J Mol Biol* 291(1):1-13.
18. Lehmann KA & Bass BL (2000) Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. *Biochemistry* 39(42):12875-12884.
19. Eggington JM, Greene T, & Bass BL (2011) Predicting sites of ADAR editing in double-stranded RNA. *Nat Commun* 2:319.
20. Wong SK, Sato S, & Lazinski DW (2001) Substrate recognition by ADAR1 and ADAR2. *RNA* 7(6):846-858.
21. Macbeth MR, *et al.* (2005) Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science* 309(5740):1534-1539.
22. Pokharel S & Beal PA (2006) High-throughput screening for functional adenosine to inosine RNA editing systems. *ACS Chem Biol* 1(12):761-765.
23. Ohman M, Kallman AM, & Bass BL (2000) In vitro analysis of the binding of ADAR2 to the pre-mRNA encoding the GluR-B R/G site. *RNA* 6(5):687-697.
24. Stephens OM, Yi-Brunozzi HY, & Beal PA (2000) Analysis of the RNA-editing reaction of ADAR2 with structural and fluorescent analogues of the GluR-B R/G editing site. *Biochemistry* 39(40):12243-12251.

25. Macbeth MR, Lingam AT, & Bass BL (2004) Evidence for auto-inhibition by the N terminus of hADAR2 and activation by dsRNA binding. *RNA* 10(10):1563-1571.
26. Klimasauskas S, Kumar S, Roberts RJ, & Cheng X (1994) HhaI methyltransferase flips its target base out of the DNA helix. *Cell* 76(2):357-369.
27. Holz B, Klimasauskas S, Serva S, & Weinhold E (1998) 2-Aminopurine as a fluorescent probe for DNA base flipping by methyltransferases. *Nucleic Acids Res* 26(4):1076-1083.
28. Allan BW, *et al.* (1999) DNA bending by EcoRI DNA methyltransferase accelerates base flipping but compromises specificity. *J Biol Chem* 274(27):19269-19275.
29. Youngblood B, Bonnist E, Dryden DT, Jones AC, & Reich NO (2008) Differential stabilization of reaction intermediates: specificity checkpoints for M.EcoRI revealed by transient fluorescence and fluorescence lifetime studies. *Nucleic Acids Res* 36(9):2917-2925.
30. Youngblood B & Reich NO (2006) Conformational transitions as determinants of specificity for the DNA methyltransferase EcoRI. *J Biol Chem* 281(37):26821-26831.
31. Cheng X & Roberts RJ (2001) AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic Acids Res* 29(18):3784-3795.
32. Lenz T, *et al.* (2007) 2-Aminopurine flipped into the active site of the adenine-specific DNA methyltransferase M.TaqI: crystal structures and time-resolved fluorescence. *J Am Chem Soc* 129(19):6240-6248.
33. Daujotyte D, *et al.* (2004) HhaI DNA methyltransferase uses the protruding Gln237 for active flipping of its target cytosine. *Structure* 12(6):1047-1055.
34. Wong I, Lundquist AJ, Bernards AS, & Mosbaugh DW (2002) Presteady-state analysis of a single catalytic turnover by Escherichia coli uracil-DNA glycosylase reveals a "pinch-pull-push" mechanism. *J Biol Chem* 277(22):19424-19432.
35. Roberts RJ & Cheng X (1998) Base flipping. *Annu Rev Biochem* 67:181-198.
36. Coman D & Russu IM (2005) A nuclear magnetic resonance investigation of the energetics of basepair opening pathways in DNA. *Biophys J* 89(5):3285-3292.
37. Krueger A, Protozanova E, & Frank-Kamenetskii MD (2006) Sequence-dependent base pair opening in DNA double helix. *Biophys J* 90(9):3091-3099.

38. Hart K, Nystrom B, Ohman M, & Nilsson L (2005) Molecular dynamics simulations and free energy calculations of base flipping in dsRNA. *RNA* 11(5):609-618.
39. Li M, *et al.* (2010) Mutational spectrum of the ADAR1 gene in dyschromatosis symmetrica hereditaria. *Arch Dermatol Res* 302(6):469-476.
40. Tojo K, *et al.* (2006) Dystonia, mental deterioration, and dyschromatosis symmetrica hereditaria in a family with ADAR1 mutation. *Movement disorders : official journal of the Movement Disorder Society* 21(9):1510-1513.
41. Kondo T, *et al.* (2008) Dyschromatosis symmetrica hereditaria associated with neurological disorders. *The Journal of dermatology* 35(10):662-666.
42. Macbeth MR & Bass BL (2007) Large-scale overexpression and purification of ADARs from *Saccharomyces cerevisiae* for biophysical and biochemical studies. *Methods Enzymol* 424:319-331.

Supplementary information

Supplementary methods

Construction of YEpTOP2PGAL1-RSinADAR2. The restriction site, *Sph*I, was introduced into YEpTOP2PGAL1 by sewing PCR using primers RSinADAR1up and RSinADAR1down, and primers RSinADAR2up and RSinADAR2down (primer sequences in Supplementary Table 2.S1). The PCR product and plasmid YEpTOP2PGAL1 were then digested with *Age*I and *Bsr*GI, followed by ligation to produce plasmid YEpTOP2PGAL1-RSinADAR2.

Construction of YCp-ADAR2 and YCp-RSinADAR. ADAR2 genes were amplified by PCR using both YEpTOP2PGAL1 and YEpTOP2PGAL1-RSinADAR2 as templates, and using primers F1-YCp ADAR2WT and R1-YCp ADAR2WT. Both PCR products and the YCp vector were digested with *Xba*I and *Eco*RI, followed by ligation to produce YCp-ADAR2 and YCp-RSinADAR2.

Construction of pR/GαGal-GAC. GAC hairpin-reporter was constructed by performing sewing PCR using primers upLEFT and upRIGHT, and primers downLEFT and downRIGHT, with pR/GαGal as template. We further digested the PCR product and plasmid with *Bst*XI and *Spe*I followed by ligation.

Cloning UAG and GAC hairpin-reporters into YIp vector. Both UAG and GAC hairpin-reporter constructs were PCR amplified using templates pR/GαGal and pR/GαGal-GAC, respectively, and primers F1-YCpαGal and R1-YCpαGal. These PCR products and YCp vector were digested with *Spe*I and *Xma*I, followed by ligation of the insert with the vector. Both YCp vectors containing UAG or GAC hairpin-reporters were digested with *Sac*I and *Kpn*I, and the resulting insert containing the *ADHI* promoter,

hairpin-reporter and *CYCI* terminator, was ligated with YIp (Yeast Integrating Plasmid) that was also digested with *SacI* and *KpnI*.

Integration of hairpin-reporter into yeast chromosomes. YIp vectors containing either UAG or GAC hairpin-reporters were linearized using restriction enzyme *MfeI*, which cleaves the plasmid in the *TRP1* gene. The linearized plasmid was transformed into W303 α yeast strain by lithium acetate transformation. The W303 α strain is auxotrophic for *TRP1*, thus if the linearized plasmid is integrated into the chromosome by homologous recombination, the yeast strain becomes prototrophic for *TRP1*, and can be selected on CM-TRP plates. We further confirmed that hairpin-reporters were integrated into the chromosome by performing colony PCR with primers YCp α GalSeqP1 and YCp α GalSeqP2.

Preparation of gapped vector and template for the screen. YCp-RSinADAR2 was digested with *SphI* and *EcoRI* to produce a gapped vector lacking the catalytic domain. ADAR2 catalytic domain was PCR amplified using YCp-ADAR2 as template and primers CDRanMutP1 and CDRanMutP3.

Cloning mutants identified in the screen into a yeast expression plasmid. Single mutations identified in the screen were subcloned from the YCp vector into YEptOP2PGAL1 using restriction enzymes *AgeI* and *PspXI*. To construct the double-mutant, N597K/N613K, we first created N597K/N613K in YCp, by performing sewing PCR using primers CDRanMutP0 and N613KP1down, and CDRanMutP3 and N613KP3up, with N597K in YCp as template. We further digested the PCR product and the plasmid, N597K in YCp, with *XmaI* and *PspXI* followed by ligation to produce N597K/N613K in YCp. Full-length N597K/N613K and truncN597K/N613K in

YE_pTOP2PGAL1 were constructed by subcloning from N597K/N613K in YCp using restriction enzymes, *Mlu*I and *Bst*EII.

RNA preparation. To prepare 5' end-labeled hairpins, chemically synthesized UAG and GAC hairpins were gel-purified after 6% denaturing PAGE, 5' end-labeled using T4 polynucleotide kinase (New England Biolabs) and γ -³²P ATP (PerkinElmer), passed through CHROMA SPIN-30 (Clontech) to remove free nucleotides, and separated by 6% denaturing PAGE. The appropriate gel band was excised, and extracted overnight in 0.5x TBE at 4 °C. RNA was ethanol precipitated, redissolved in refolding buffer (50mM HEPES (pH 7.5), 5mM EDTA, 250mM KCl), heated to 95 °C for 4 min, slow cooled to room temperature for 2 hr, and again gel-purified from a 6% native polyacrylamide gel.

To prepare internally radiolabeled hairpins, 20 and 22 nt DNA:RNA chimeras were separated by 20% denaturing PAGE, 53 and 54 nt RNA and 73 and 76 nt DNA were separated by 6% denaturing PAGE, and gel-purified. For UAG hairpin, 100 pmoles of a 54 nt RNA (5'-AGGUGGGUGGAAUAGUAUACCAUUCGUGGUAUAGUAUCCCA CCUACCCAGACGG-3') was 5' end labeled as above, passed through CHROMA SPIN-10 (Clontech) to remove free nucleotides, extracted with phenol/chloroform/isoamyl alcohol, and ethanol precipitated. To this RNA, we added 200 pmoles of a 22 nt DNA:RNA chimera, d(CCAGGCGTTTTTGGGT)r(CCGUUU), where the nucleotides following "d" are deoxyribonucleotides and those following "r" are ribonucleotides. To the above mixture we added 120 pmoles of a 76 nt DNA splint (5'-CCGTCTGGGTAGG TGGGATACTATACCACGAATGGTATACTATTCCACCCACCTAAACGGACCCA AAAACGCCTGG-3'), volume was brought to 20 μ l with double distilled (dd) water, and the mixture heated to 95 °C for 4 min and slow cooled to room temperature (RT) for

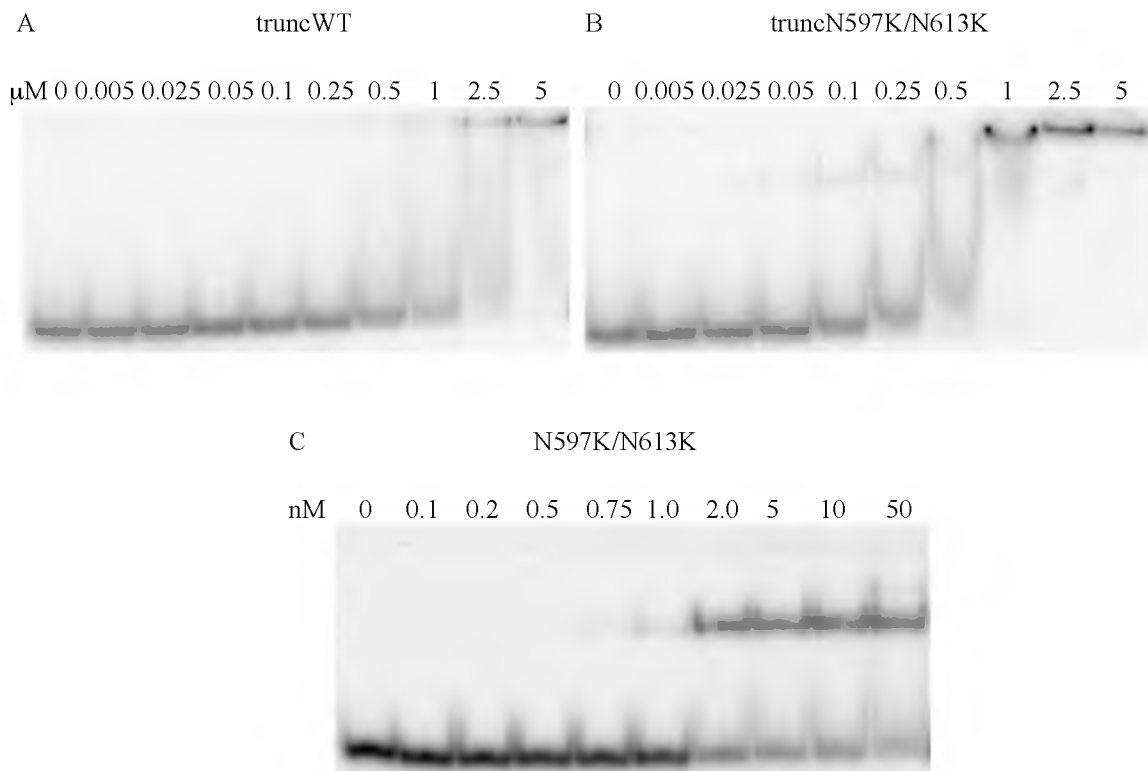
2 hr. To this hybridized mixture, we added 4 μ l of 10x T4 DNA ligase buffer (500 mM Tris-HCl (pH 7.5), 100 mM MgCl₂, 100 mM DTT, 10 mM ATP, 250 μ g/ml BSA), 4 μ l of 10 mM ATP, 2 μ l RNasin, 6 μ l of T4 DNA ligase and 4 μ l of dd water to bring the volume to 40 μ l, and then incubated the mixture at room temperature overnight. Then 2 μ l of RQ1 RNase free DNase was added, followed by incubation at 37 °C for 30 min, and products were purified from 6% denaturing polyacrylamide gels. Gel bands were excised, extracted overnight in 0.5x TBE at 4 °C, ethanol precipitated, and RNAs refolded as described above. To prepare the internally labeled GAC hairpin, we used a 53 nt RNA (5'-ACUGGGUGGAAUAGUAUACCACUCACGGUAUAGUAUCCACCU AGCCGGACGG-3'), a 20 nt DNA:RNA chimera, d(GGCGTTTTTGGGT)r(CCGUUU G), and a 73 nt DNA splint (5'-CCGTCCGGCTAGGTGGGATACTATACCGTGAGT GGTATACTATTCCACCCAGTCAAACGGACCCAAAAACGCC-3').

To prepare the 418 bp RNA, sense and antisense strands were synthesized from two separate PCR products using T7 RNA polymerase. The PCR products were amplified using pSP65 CAT A plasmid as template. pSP65 CAT A plasmid has been described (1). Primers 10 and 4 were used to amplify the sense PCR product, and primers 1 and 3 were used to amplify the antisense PCR product (primer sequences in Supplementary Table I). For synthesis of internally radiolabeled 418 bp RNA, α -³²P ATP (PerkinElmer) was included in the transcription reaction. Equal amounts of sense and antisense strands were dissolved in hybridization buffer (10 mM Tris-HCl (pH 8), 40 mM KCl), heated to 95 °C for 4 min, slow cooled to room temperature for 2 hr, ethanol precipitated and purified from a 4.5% native polyacrylamide gel.

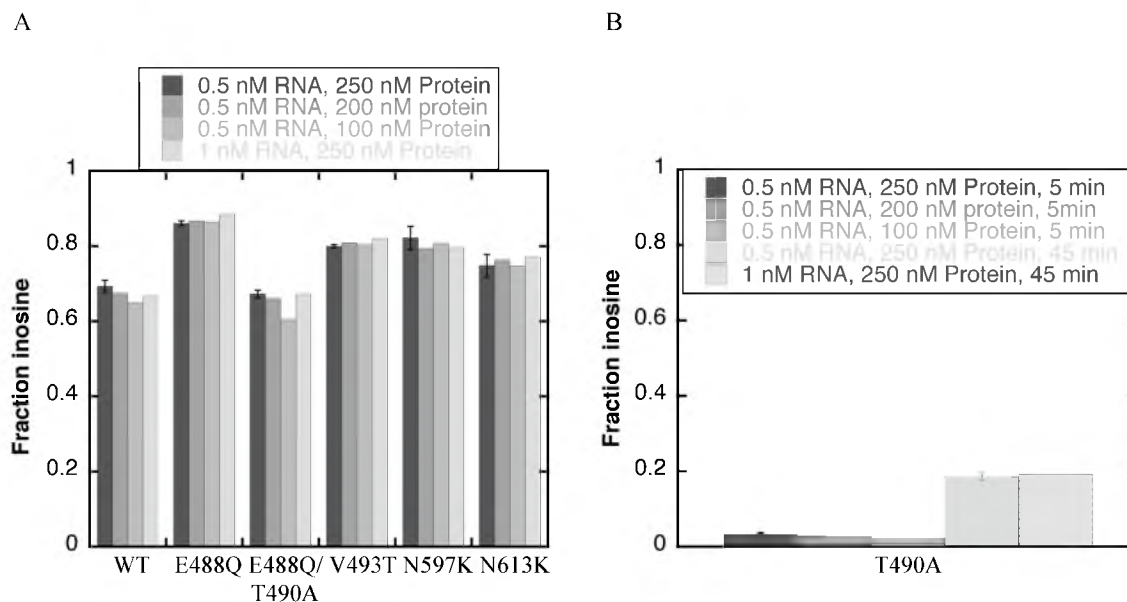
Preference assay. For performing the initial time course with internally

radiolabeled 418 bp dsRNA, 0.25 nM RNA and 250 nM protein were incubated at 20 °C for: 0 sec, 20 sec, 30 sec, 1 min, 5 min, 10 min, 20 min, 30 min, 60 min and 90 min. The buffer was that used for deamination assays. The time required to achieve ~ 20% editing was determined, and preference assays were performed for the same duration. Editing reactions were stopped by addition of 0.5% SDS, heating to 95 °C for 1 min and placed on ice. Proteinase K buffer was added (10x stock = 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 5 mM EDTA, 0.5 % SDS and 3.33 µg/ µl Proteinase K), incubation was continued at 37 °C for 20 min followed by RNA extraction and ethanol precipitation. cDNA was made from purified RNA with reverse transcriptase (Thermoscript, invitrogen) using primer 6 for sense strands and primer 5 for antisense strands. RNA was removed by treating with RNase H, and cDNA was PCR amplified using primers 6 and 12 (sense strands), or primers 5 and 7 (antisense strands). Sense and antisense PCR products were sequenced with primer 6 and primer 5, respectively (GENEWIZ), and editing sites quantified as described (2).

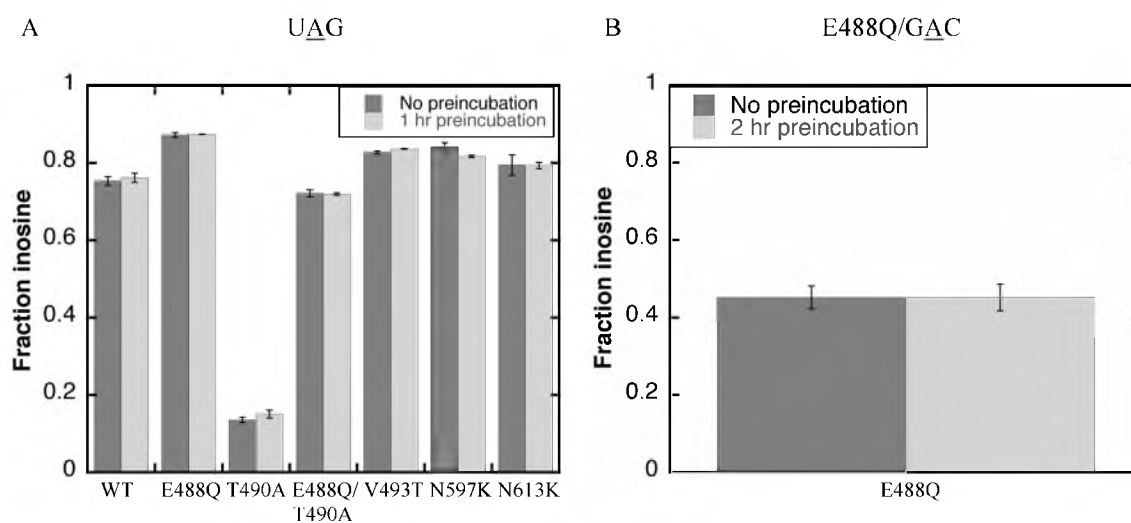
Supplementary results



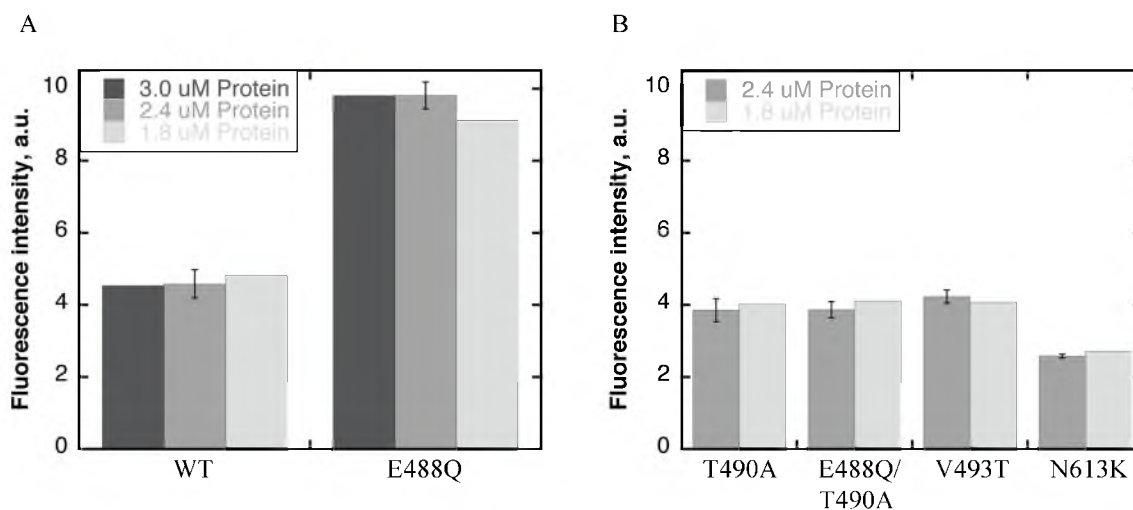
Supplementary Fig. 2.S1. **Binding affinity of truncN597K/N613K is ~4-fold higher than truncWT.** (A and B) PhosphorImages showing representative gel shift assays of truncWT and truncN597K/N613K, respectively, with 20 pM of ³²P-5' end-labeled UAG hairpin. The dissociation constant, K_d , for truncWT and truncN597K/N613K is 527.5 ± 14.9 nM and 148.1 ± 1.9 nM, respectively. Protein concentrations (μM) are indicated at the top of each gel. Error = s.d. ($n \geq 2$). (C) PhosphorImage showing gel shift for full-length double-mutant N597K/N613K performed as a control with UAG hairpin. Dissociation constant of the double-mutant, N597K/N613K (~1.3 nM) is similar to that of the single mutants, N597K and N613K.



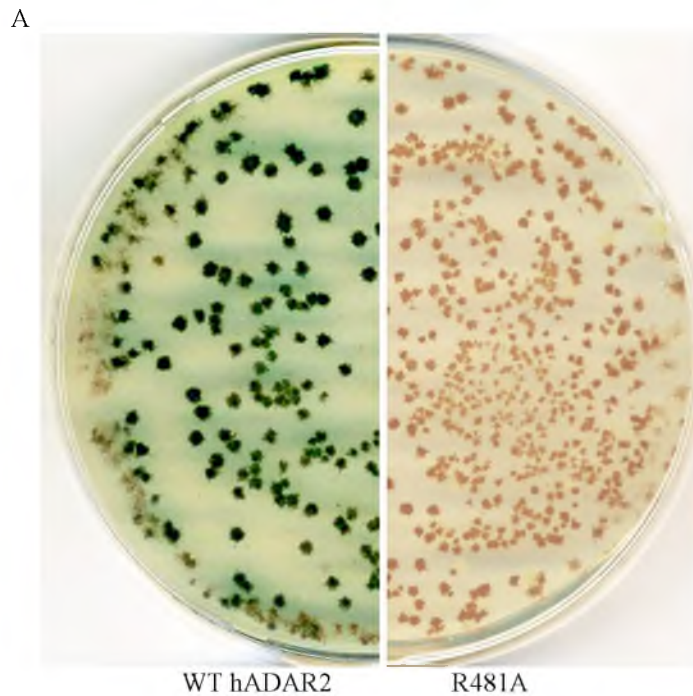
Supplementary Fig. 2.S2. **Experiments to test single turnover kinetics.** (A) Bar graph depicting fraction inosine produced for each protein with UAG hairpin, using either less protein or more RNA. Measurements were made for the 5 min timepoint. Error bar = s.d. ($n \geq 3$) for measurements made using 0.5 nM RNA and 250 nM protein, which was used for determining the deamination rate, k_{deam} . (B) Bar graph depicting fraction inosine produced with T490A mutant as in (A). All measurements were made at 5 min timepoint, except measurements with 1 nM RNA and 250 nM protein which were made at 45 min time point. For comparison, the fraction inosine produced at 45 min time point using 0.5 nM RNA and 250 nM protein (concentrations used for determining deamination rate, k_{deam}) is also plotted. Error bar for this concentration = s.d. ($n \geq 3$).



Supplementary Fig. 2.S3. **Confirmation that enzymes retain activity for the course of the experiment.** (A) Bar graph depicting fraction inosine produced with WT hADAR2 and mutants without, and with, preincubation for 1 hr at 20 °C prior to incubation with UAG hairpin. Incubation was for 10 min in case of WT hADAR2, E488Q, E488Q/T490A, V493T, N597K and N613K, and for 30 min with T490A. Error bar = s.d. ($n \geq 3$). (B) Bar graph depicting fraction inosine produced for E488Q without, and with, preincubation for 2 hr at 30 °C prior to incubation with GAC hairpin for 40 min. Error bar = s.d. ($n \geq 3$).



Supplementary Fig. 2.S4. **Control experiments with UA₂APG-28 to determine saturating protein concentrations for the 2-AP fluorescence assay.** (A) Bar graph depicting increase in FI, in arbitrary units, on addition of 3.0, 2.4 and 1.8 μ M WT hADAR2 or E488Q to 0.6 μ M UA₂APG-28 duplex. 2.4 μ M protein concentration was selected for the 2-AP fluorescence assays, and for this concentration, error bar = s.d. ($n \geq 3$). (B) For other mutants, bar graph is plotted depicting increase in FI, in arbitrary units, on addition of 2.4 and 1.8 μ M protein to 0.6 μ M UA₂APG-28 duplex. For the 2.4 μ M protein concentrations, error bar = s.d. ($n \geq 3$). Increase in FI with 1.8 μ M protein was similar to that determined with 2.4 μ M protein.



Supplementary Fig. 2.S5. ***In vivo* α -galactosidase assay to test editing of UAG with R481A mutant.** (A) CM-URA plates with yeast colonies that have UAG hairpin-reporter integrated into the chromosome, and transformed with R481A mutant. WT hADAR2 is shown as a control.

Supplementary Table 2.S1. **Primer sequences**

Primer name	Primer sequences	Use
RSinADAR1up	TCGTCTTCACCGGTCGCGTTCCTGAAACG	hADAR2 plasmid
RSinADAR1down	TCAGCTAAACCTGCGGTAAATGCAGGCATGCCTGAAGACCCTCACTGGGAATAGGCTGG	
RSinADAR2up	GCATGCCTGCATTTACCGCAGGTTTTAGCTGACGCTGTCTCACG	
RSinADAR2down	GGTGCTGATGTACAGATGAAACTGG	
F1-YCpADAR2WT	CGGATTCTAGAATGTACACCATCACCATCACC	
R1-YCpADAR2WT	GATATCGAATTCTCAGGGCGTGAGTGAGAACTGG	
CDRanMutP1	AGTCTGCCCTGGCCGCCATTTTAACTTGC	
CDRanMutP3	ACCTAGACTTCAGGTTGTCTAACTCC	
CDRanMutP0	CCAAGGCCCGGGCTGCGCAGTCTGCC	
N613KP1down	AGTCGTGGCCTTGATGACCTCAATAGCG	
N613KP3up	GAGGTCATCAAGGCCACGACTGGG	
upLEFT	GGATGATCCACTAGTACGG	Hairpin-reporter
upRIGHT	GATACTATACCGTGAGTGGTATACTATTCCACCCAGTCA AACGGACCCAAAAACGCCTGGC	
downLEFT	TAGTATACCACTCACGGTATAGTATCCCACCTAGCCGGA CGGGATCCCCGAGTTACAATGGCCTTGGTCTCACTCC	
downRIGHT	TCTGCAACATGGCCCATACC	
F1-YCp α Gal	CTCTAGAAGTAGTATGAGAGCTTTCTTGTTCACCG	
R1-YCp α Gal	TTCCTGCAGCCCCGGGTCAAGAAGAGGGTCTCAACC	
YCp α GalSeqP1	TGTTTCCTCGTCATTGTTCTCG	
YCp α GalSeqP2	ACTCCTTCCTTTTCGGTTAGAGC	
488P1down	CGCATTTGGAGCGCACTGGAATCGTCCNNNACCAGACT CTATTTTGGTCCG	
488P3up	GGGACGATTCCAGTGCGCTCCAATGCG	Mutational analysis
490P1down	GGATGCTCGCATTGGAGCGCACTGGAATNNNCCCCTCA CCAGACTCTATTTTG	
490P3up	ATTCCAGTGCGCTCCAATGCGAGCATCC	
493P1down	CCCACGTTTGGATGCTCGCATTGGAGCGNNNTGGAATC GTCCCCTCACC	
493P3up	CGCTCCAATGCGAGCATCCAAACGTGGG	
597P1down	GCCTACCGTCCAGTTGACACTGAANNNGGGGCCTTCC CTGGC	
597P3up	TTCAGTGTCAACTGGACGGTAGGC	
613P1down	CCCAGCTCATCCTTCCCAGTCGTGGCNNNGATGACCTCA ATAGCG	
613P3up	GCCACGACTGGGAAGGATGAGCTGGG	
10	ATCGAGTCTATAATACGACTCACTATAGTAGCCTTGAGC TTGGATCTGCCAGCTTGGCGAGATTTTCAGG	418 bp dsRNA
4	TTAGGAGCAACGAACACGCCACATCTTGCG	
1	CTCGTACAATTAATACGACTCACTATAGATAGGCCAGGT TTTACCG	
3	TTCAGGTTTGGATCCCAGCTTGGCGAGATTTTCAGG	
6	TTAGGAGCAACGAACACGC	Preference assay
12	GTAGCCTTGAGCTTGGATCTGCC	
5	TTCAGGTTTGGATCCCAGC	
7	GATAGGCCAGGTTTTTCACCG	

References

1. Harland R & Weintraub H (1985) Translation of mRNA injected into *Xenopus* oocytes is specifically inhibited by antisense RNA. *J Cell Biol* 101(3):1094-1099.
2. Eggington JM, Greene T, & Bass BL (2011) Predicting sites of ADAR editing in double-stranded RNA. *Nat Commun* 2:319.

CHAPTER 3

PERSPECTIVES

Summary

Although it has been known for more than a decade that the extent to which an adenosine is edited depends on its 5' and 3' nearest neighbors, the residues in ADAR that are involved in preferences and the underlying mechanism have long been outstanding questions. The studies in this dissertation have identified for the first time, residues in hADAR2 that affect preferences, and point to a conserved loop close to the active site as being of key importance. Unexpectedly, the studies also suggest that preferences are derived from differences in base-flipping rather than direct recognition of the neighboring bases. In this section, I will discuss the outstanding questions and the experiments that can be performed to address these questions. Additionally, I will discuss the findings in this dissertation in the context of previous studies of ADARs and other deaminases, in hopes of highlighting how they further our understanding of enzyme mechanism.

Residues in ADAR involved in base-flipping and preferences

The screen described in this dissertation identified eight mutations in the hADAR2 catalytic domain that affect preferences. Three mutations, E488Q, T490A and V493A were on a highly conserved loop that includes two β -strands (shown in green, Fig. 2.1F).

The E488Q mutant showed an increase in 2-AP FI with both UA_{2AP}G-28 and GA_{2AP}C-27 duplexes, which is suggestive of an increase in base-flipping, and a corresponding increase in deamination rate as well. Compared to other mutations characterized, the E488Q mutant showed the most dramatic increase in editing in both our *in vivo* α -galactosidase reporter assays as well as in the *in vitro* single turnover catalytic rate determinations, and our data suggest that this increase is due to enhanced base-flipping. Therefore, the next obvious questions are: (a) how does a Gln residue at this position enhance base-flipping? (b) does this residue facilitate base-flipping directly, or indirectly via interactions with other residues? (c) in the wildtype enzyme does a Glu residue at 488 mediate base-flipping? Another base-flipping enzyme, cytosine-5 DNA methyltransferase M.HhaI, utilizes a Gln (Q237) to flip out the target cytosine by directly pushing this cytosine out of the helix (Fig. 3.1) (1). This Gln is flanked by Gly residues, proposed to be crucial for positioning the Gln side chain for deep penetration into the helix. E488Q also has flanking Gly residues, possibly for similar penetration into the helix. Further, looking at the position of E488 in the crystal structure of hADAR2 catalytic domain with the modeled in AMP (Fig. 2.1F) (2), it is likely that E488Q directly flips out the target adenosine. However, an indirect interaction with other residues leading to base-flipping cannot be ruled out in the absence of a co-crystal with RNA. Since a Glu to Gln mutation at residue 488 enhances base-flipping, it is very likely that this residue is also involved in base-flipping in the wildtype enzyme. However, various approaches can be utilized to directly address these questions, as well as further our understanding of the residues involved in preferences. Some of these approaches are discussed below.

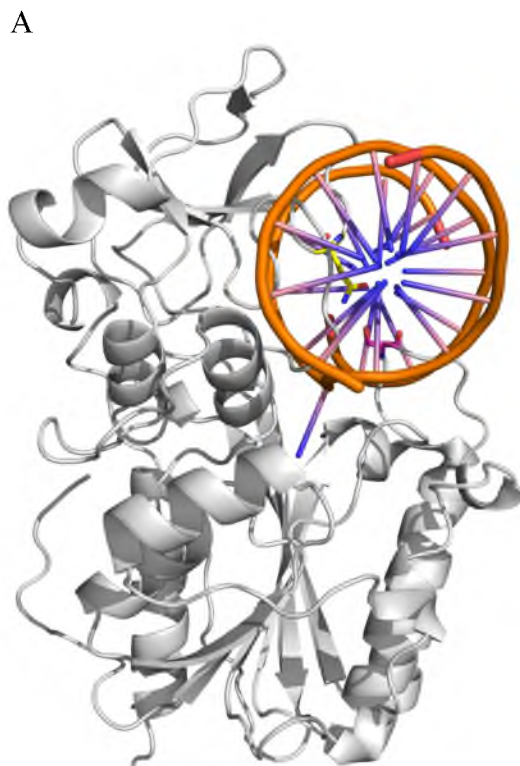


Figure 3.1. **Crystal structure of M. HhaI.** (A) Crystal structure of M. HhaI in complex with a 13-mer DNA duplex containing the recognition sequence (PDB code: 1MHT) (1). Q237 (yellow stick) pushes the target cytosine out of the helix, and S87 from the catalytic loop (pink stick) hydrogen bonds with Q237 to lock the cytosine in a flipped out position.

Co-crystallization of ADAR2 with RNA

Solving a co-crystal of ADAR2 with RNA will be the most conclusive way to gain evidence for a base-flipping mechanism by this enzyme, and will also aid in identifying or predicting residues that are involved in base-flipping. Co-crystal structures have proven base-flipping for various DNA methyl transferases and DNA glycosylases. Further, solving a co-crystal of ADAR2 with both favored UAG and disfavored GAC duplexes will be very helpful for understanding the mechanistic details of ADAR's preferences. Until now, a crystal structure has only been solved for the catalytic domain of hADAR2, and the catalytic domain alone has very weak affinity for RNA. Two of the mutants studied in this dissertation, N597K and N613K, have 2-fold higher binding affinity than WT hADAR2 for both favored UAG and disfavored GAC hairpins. Further, when binding affinities were determined for truncations comprised of only the catalytic domains, truncWT and truncN597K/N613K, truncN597K/N613K showed ~4 fold higher binding affinity compared to truncWT. Therefore, solving a co-crystal structure with truncN597K/N613K is more likely to succeed than with truncWT. Further, as mentioned in Chapter 2, if E488Q is stabilizing the flipped out adenosine, then this mutant is also more likely to yield a co-crystal than the WT enzyme. In addition, replacing the target adenosine with a nonhydrolyzable analog, nebularine, could trap the flipped out intermediate (3). An alkyl disulfide tether has also been used in some studies to obtain co-crystals (4).

Time-resolved fluorescence measurements

Time-resolved fluorescence measurements can be made with the same RNA duplexes as used in the steady-state fluorescence measurements, where the target adenosine is replaced with 2-AP. 2-AP shows a characteristic fluorescence decay on flipping out of the double helix, that involves loss of the short decay component (~100 ps) and increase in amplitude of the long decay component (~10 ns). Our 2-AP fluorescence measurements suggest that E488Q enhances base-flipping of the target adenosine. However, from the steady state measurements made, it is difficult to determine if E488Q is stabilizing the flipped out adenosine, or alternatively, if it is increasing the rate of flipping. Performing time-resolved fluorescence measurements will be crucial to distinguish between these two possibilities. Further, our data also suggest that the 2-AP FI observed in our steady-state experiments reflects both decrease in FI due to binding and an increase in FI due to base-flipping. Analysis of time-resolved fluorescence measurements will allow the separation of the two components of binding and base-flipping, and give a more accurate estimate of base-flipping.

Mutational analysis of residues on the loop

Our studies showed that three residues, E488, T490 and V493, on a conserved loop (shown in green, Fig. 2.1F) close to the active site are important for preferences. Of particular interest is residue E488, since our data suggests that this residue is involved in base-flipping. Our mutational analysis of residue 488 using the *in vivo* α -galactosidase reporter assay showed that an alanine at this position cannot edit the disfavored GAC, but it can still edit the favored UAG. Further, an asparagine at position 488 can edit both

UAG and GAC, similar to a glutamine at this position. However, the α -galactosidase reporter assay (X- α -Gal plate assay) used for our screen is not sensitive enough to discern 5-10 fold differences in catalytic rates. Hence, these variants must be purified and their catalytic rates and base-flipping determined. For example, E488A and E488D could be tested to determine if a shorter side-chain will impede base-flipping, using the 2-AP fluorescence assay, and if the decreased base-flipping will correlate with reduced activity. Although, we identified several mutations at residue 488 that allowed editing of GAC and/or UAG using the *in vivo* α -galactosidase reporter assay, making additional mutations at this position that were not initially identified in the screen will be informative. Mutating other residues on this conserved loop will be informative as well. For example, mutating the Gly residues flanking E488 into a less flexible amino acid will be helpful in determining if these Gly residues are required for penetration of E488 into the helix.

Is the conformational stability of the conserved loop essential for efficient base-flipping? In both the *in vivo* α -galactosidase reporter assay as well as in the *in vitro* determination of catalytic rate, we found that T490A did not edit adenosine in the context of GAC, and poorly edited adenosine in the context of UAG. Correspondingly, the T490A mutant also showed minimal base-flipping, as measured by the 2-AP fluorescence assay, which was not enhanced in the double mutant, E488Q/T490A. Thus, our data suggest that T490 is required for efficient base-flipping, and our favored model is that this residue is important for maintaining a conformation of the conserved loop that promotes base-flipping. Using a tryptophan fluorescence assay, previous studies with hADAR2 have shown that the conformation of the protein changes on binding RNA, in

particular that of the catalytic domain, since all five tryptophans in hADAR2 are in the catalytic domain (5). Thus, a similar tryptophan fluorescence assay could be used to determine if the observed conformational change varies between WT hADAR2 and the T490A mutant. Additionally, mutational analysis could also be done to gain indirect evidence for the importance of the conformational stability of the conserved loop. The crystal structure of hADAR2 catalytic domain shows two hydrogen bonds from the backbone carbonyl and side chain hydroxyl of T490 to the side chain of R481. Consistent with our hypothesis that T490 is required for the conformational stability of this conserved loop, we found that the R481A mutant is completely inactive in the *in vivo* α -galactosidase reporter assay. Further, T482 also forms a hydrogen bond from the side chain hydroxyl to the backbone carbonyl of another residue, I491, in this loop. Residues K483 and E485 in the conserved loop also form hydrogen bonds from their side chains to other residues in the protein. Thus, it would be informative to mutate these residues to understand their functions in the conserved loop.

The majority of the other mutants identified from the screen were on another loop close to the active site (shown in blue, Fig. 2.1F, Fig. 3.2A). This loop is part of a basic patch on the enzyme surface observed in the crystal structure of the hADAR2 catalytic domain (Fig. 1.3C), suggesting this loop is involved in binding to RNA. Therefore, it will be informative to mutate other residues on this loop that were not identified in the screen to determine if any of these residues have a role in preferences. Of particular interest are residues R590 and K594 that protrude towards the modeled in AMP (Fig. 3.2). It is likely that RNA binds to the groove between the two loops (shown in blue and green, Fig. 3.2).

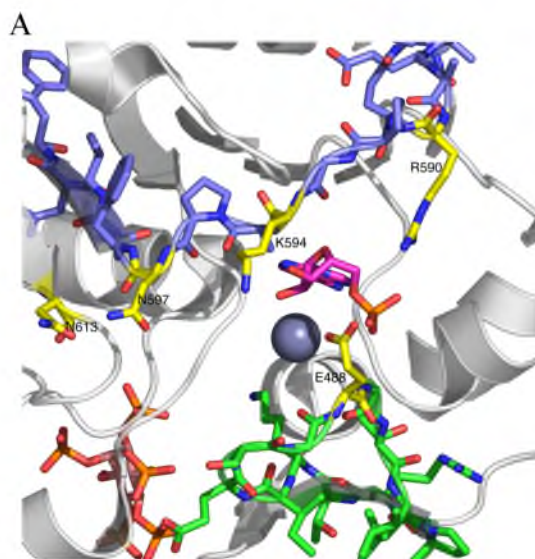


Figure 3.2. **Loops involved in preferences.** (A) Close view of two loops involved in preferences, which are in close proximity to the active site (PBD code: 1ZY7) (2). Important residues on the loops (yellow sticks), Zn (grey sphere) and modeled in AMP (pink stick) are shown.

ADAR1 has the same 5' nearest neighbor preferences as ADAR2, and similar but distinct 3' nearest neighbor preferences. Residues N597 and N613 are conserved in ADAR2 from different species, however they are negatively charged residues in ADAR1. Hence, it will be interesting to mutate N597 and N613 into negatively charged residues in ADAR2 and determine if these residues are responsible for the slight difference in preferences between ADAR1 and ADAR2.

Further base-flipping and catalytic rates must be measured with mutated RNA hairpins that have the target adenosine in context of other triplets, to determine if our theory that preference is derived from differential base-flipping holds true for other triplets as well.

Crosslinking using 5-Iodouridine

The residues in close proximity to the 5' neighbor potentially can be identified by substituting the uridine 5' to the target adenosine with 5-Iodouridine, which is a photoreactive chromophore that covalently crosslinks with protein residues upon irradiation with UV light. 5-Iodouridine has been used in various studies to identify protein residues interacting with substrate DNA or RNA (6-9). Most of the studies have identified aromatic residues like tyrosine or phenylalanine as well as basic residues like arginine or histidine crosslinked to 5-Iodouridine (10). Preliminary crosslinking experiments were performed using UV light at 308 nm wavelength, and under these conditions the crosslinked product obtained was <1% of the total RNA. Low crosslinking yield is a major problem with this technique. Previous studies have shown that 325 nm monochromatic UV light is optimal for crosslinking using 5-Iodouridine, and minimizes nonspecific interactions. Therefore, irradiating at 325 nm wavelength using a HeCd laser might improve the yield. However, if 5' uridine is not in close proximity to protein residues, then very little or no crosslinked products might be observed.

How are editing sites specifically selected by other deaminases?

As mentioned in Chapter 1, ADA and CDA deaminate free nucleotides, and hence will not have a substrate RNA recognition mechanism like ADARs. However, ADATs perform A-to-I deaminations in tRNA, and a co-crystal of a prokaryotic TadA, which belongs to the ADAT2 family, in complex with tRNA^{Arg2} revealed that a loop following β -strand 4 (shown in orange, Fig 3.3A) is close to the active site and contacts the nucleotide upstream of the target adenosine (3, 11). APOBECs also show preferences for

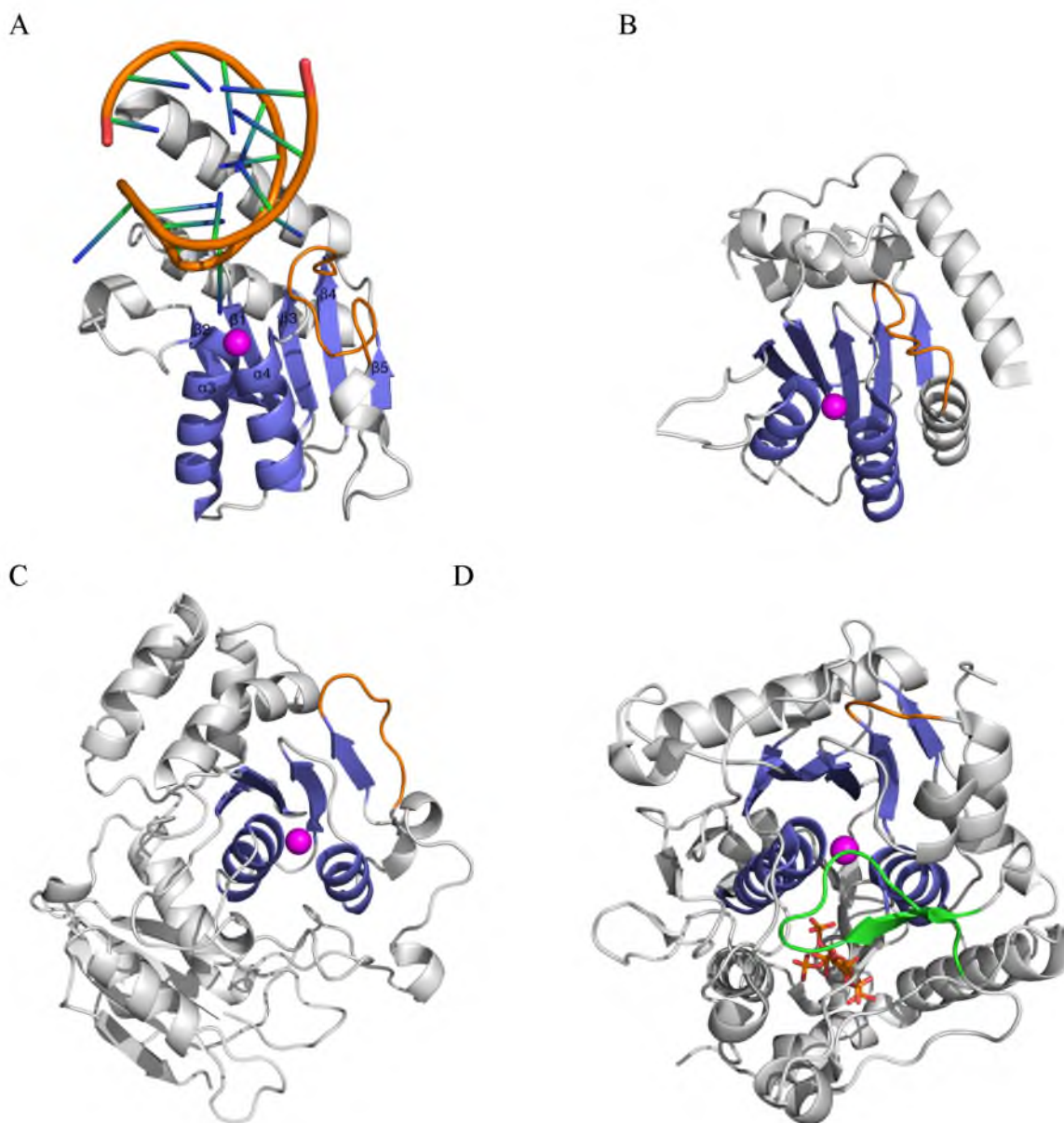


Figure 3.3. Crystal structures of deaminases showing loops involved in specificity. (A, B, C and D) Crystal structures of *S. aureus* TadA in complex with the anticodon stem-loop of tRNA^{Arg2}, APOBEC3G catalytic domain from *E. coli*, CDA from *E. coli*, and hADAR2 catalytic domain respectively (PDB code: 2B3J, 3E1U, 1CTU, and 1ZY7 respectively) (2, 3, 12, 13). The conserved deaminase motif (colored blue), loop involved in specificity in TadA and APOBEC3 (colored orange) (3, 11) and loop involved in specificity in hADAR2 (colored green) is shown.

a 5' flanking nucleotide, and mutational studies in AID and APOBEC3 have revealed residues that are involved in specificity. Interestingly, a region that is involved in specificity in both AID and APOBEC3 also maps to a loop following β -strand 4 in the crystal structures of APOBEC2 and catalytic domain of APOBEC3G (shown in orange, Fig 3.3B) (11). However, the loop following β -strand 4 in CDA is further away from the active site (shown in orange, Fig 3.3C) (11), consistent with the fact that it acts on free nucleotides, and this loop might not be performing a similar function as in TadA and APOBEC3.

The crystal structure of ADAR2 catalytic domain also shows this loop to be further away from the active site (shown in orange, Fig 3.3D), indicating this loop may not be involved in ADAR2's 5' neighbor preferences. Studies using an eight-term model have shown that although the 5' neighbor of adenosine is most influential in determining the extent of editing, bases further away from the nearest neighbor also contribute slightly (14). This could be explained by our model that preferences are derived from differential base-flipping, and base-flipping in turn could be affected by neighboring bases. NMR experiments and theoretical calculations indicate base pair opening probability is affected by neighboring bases (15, 16). However, it will be interesting to analyze this loop (residues 548-551, GSLY) to determine if it contributes to preferences as well.

Our studies identified another conserved loop on the surface of the protein and close to the active site as important for preferences. Interestingly, this loop is absent in TadA as well as APOBECs, consistent with the fact that these enzymes do not have to flip the target adenosine to access it, since TadA acts on the tRNA anticodon loop, while APOBECs act on ssRNA or ssDNA.

Are ADARs' preferences derived from differential base flipping?

WT hADAR2 and all hADAR2 mutants tested, flipped adenosine in the context of the favored UAG more than in the context of the disfavored GAC, as indicated by an increase in 2-AP FI on addition of protein to RNA that has the target adenosine replaced with 2-AP. Base-flipping for all proteins correlated with their respective deamination rates. Additionally, an ADAR2 mutant, E488Q, that showed an increased catalytic rate with both UAG and GAC, also showed an increase in base-flipping with each of them. Thus, whether it is a comparison of the WT protein and mutants, or favored substrate and disfavored substrate, in all our experiments, increased catalytic rate correlated with an increase in base-flipping, as indicated by an increase in 2-AP FI, suggesting that ADARs preferences are derived from differential base-flipping.

Are there other enzymes that affect specificity by modulating base-flipping? In M.EcoRI, an N6-adenine DNA methyl transferase that utilizes a bending, base-flipping and intercalation mechanism, a bending deficient mutant decreases base-flipping and increases specificity (17, 18). For noncognate substrates, M.EcoRI specificity arises from partitioning the enzyme/DNA intermediate into the unbent form (18, 19).

ADARs are proposed to have evolved from CDA, however, studies have also shown that based on sequence alignments, ADARs are also similar to cytosine-5 DNA methyl transferases and N6-adenine DNA methyl transferase (20). Thus, it is possible that ADARs utilize a base-flipping mechanism similar to that of methyl transferases, and differential base-flipping is responsible for preferences.

What is the rate-limiting step of the deamination reaction?

Slow turnover rate and substrate inhibition of ADARs make steady state measurements challenging (21). Yet, steady state rate measurements using hADAR2 and a large excess of a favored substrate (substrate derived from the R/G site of GRIA2 pre-mRNA) showed that product formation is linear with time upto three turnovers with no burst in product formation (21). This indicates that product release is not the rate-limiting step (21). Additionally, it has been proposed that binding and base-flipping are most likely not the rate-limiting steps (21-23). Further, studies with substrate analogs suggest that the chemical step is rate limiting (22, 23). Since these measurements were made using a favored substrate, it is likely that for a disfavored substrate, the rate-limiting step may be different.

With E488Q, a small increase in base-flipping of adenosine within GAC (FI ~2.6) resulted in ~60 fold increase in editing, whereas a large increase in base-flipping of the adenosine within UAG (FI ~9.8) resulted in only ~7 fold increase in editing. This can be explained, if we assume that for the disfavored substrate, GAC, the rate-limiting step is base-flipping. This is consistent with the fact that neither WT hADAR2 nor mutants could flip adenosine within GAC efficiently. The rate of flipping for favored versus disfavored substrates can be determined by performing a time-resolved fluorescence measurement for each substrate (17, 18), and once the base-flipping rate with each substrate is known, by comparison with rates for the chemical step, the rate-limiting step can be determined (17, 18).

Conclusion

It has been known for a long time that the extent to which an adenosine is edited depends on the sequence context of the target adenosine. However, the residues in ADAR that are involved in preferences and the underlying mechanism have long been an outstanding question. The studies in this dissertation have identified for the first time, residues in hADAR2 that affect preferences, and point to a conserved loop close to the active site as being of key importance. Our studies also suggest that preferences are derived from differences in base-flipping rather than direct recognition of the neighboring bases. However, many questions still remain regarding the mechanism by which ADARs achieve specificity. How do residues on the conserved loop interact with the RNA? Do ADARs indeed use a base-flipping mechanism to gain access to a base? If so, what base-flipping mechanism do ADARs use? Do ADARs actively push the base out of the helix or do they use a passive mechanism? What are the initial events that lead to base-flipping? Does the conserved loop have a similar function in ADAR1? What is the cause for aberrant A-to-I editing levels observed in diseases? Answering these questions will be very helpful in designing drugs for specific conditions.

References

1. Klimasauskas S, Kumar S, Roberts RJ, & Cheng X (1994) HhaI methyltransferase flips its target base out of the DNA helix. *Cell* 76(2):357-369.
2. Macbeth MR, *et al.* (2005) Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science* 309(5740):1534-1539.
3. Losey HC, Ruthenburg AJ, & Verdine GL (2006) Crystal structure of *Staphylococcus aureus* tRNA adenosine deaminase TadA in complex with RNA. *Nat Struct Mol Biol* 13(2):153-159.

4. Banerjee A, Santos WL, & Verdine GL (2006) Structure of a DNA glycosylase searching for lesions. *Science* 311(5764):1153-1157.
5. Yi-Brunozzi HY, Stephens OM, & Beal PA (2001) Conformational changes that occur during an RNA-editing adenosine deamination reaction. *J Biol Chem* 276(41):37827-37833.
6. Jeltsch A, Roth M, & Friedrich T (1999) Mutational analysis of target base flipping by the EcoRV adenine-N6 DNA methyltransferase. *J Mol Biol* 285(3):1121-1130.
7. Willis MC, Hicke BJ, Uhlenbeck OC, Cech TR, & Koch TH (1993) Photocrosslinking of 5-iodouracil-substituted RNA and DNA to proteins. *Science* 262(5137):1255-1257.
8. Hicke BJ, Willis MC, Koch TH, & Cech TR (1994) Telomeric protein-DNA point contacts identified by photo-cross-linking using 5-bromodeoxyuridine. *Biochemistry* 33(11):3364-3373.
9. Stump WT & Hall KB (1995) Crosslinking of an iodo-uridine-RNA hairpin to a single site on the human U1A N-terminal RNA binding domain. *RNA* 1(1):55-63.
10. Daujotyte D & Klimasauskas S (2000) Affinity photo-crosslinking study of the DNA base flipping pathway by HhaI methyltransferase. *Nucleic acids symposium series* (44):271-272.
11. Conticello SG, Langlois MA, & Neuberger MS (2007) Insights into DNA deaminases. *Nat Struct Mol Biol* 14(1):7-9.
12. Holden LG, *et al.* (2008) Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature* 456(7218):121-124.
13. Xiang S, Short SA, Wolfenden R, & Carter CW, Jr. (1995) Transition-state selectivity for a single hydroxyl group during catalysis by cytidine deaminase. *Biochemistry* 34(14):4516-4523.
14. Eggington JM, Greene T, & Bass BL (2011) Predicting sites of ADAR editing in double-stranded RNA. *Nat Commun* 2:319.
15. Coman D & Russu IM (2005) A nuclear magnetic resonance investigation of the energetics of basepair opening pathways in DNA. *Biophys J* 89(5):3285-3292.
16. Krueger A, Protozanova E, & Frank-Kamenetskii MD (2006) Sequence-dependent base pair opening in DNA double helix. *Biophys J* 90(9):3091-3099.

17. Allan BW, *et al.* (1999) DNA bending by EcoRI DNA methyltransferase accelerates base flipping but compromises specificity. *J Biol Chem* 274(27):19269-19275.
18. Youngblood B, Bonnist E, Dryden DT, Jones AC, & Reich NO (2008) Differential stabilization of reaction intermediates: specificity checkpoints for M.EcoRI revealed by transient fluorescence and fluorescence lifetime studies. *Nucleic Acids Res* 36(9):2917-2925.
19. Youngblood B & Reich NO (2006) Conformational transitions as determinants of specificity for the DNA methyltransferase EcoRI. *J Biol Chem* 281(37):26821-26831.
20. Hough RF & Bass BL (1997) Analysis of *Xenopus* dsRNA adenosine deaminase cDNAs reveals similarities to DNA methyltransferases. *RNA* 3(4):356-370.
21. Stephens OM, Yi-Brunozzi HY, & Beal PA (2000) Analysis of the RNA-editing reaction of ADAR2 with structural and fluorescent analogues of the GluR-B R/G editing site. *Biochemistry* 39(40):12243-12251.
22. Maydanovich O & Beal PA (2006) Breaking the central dogma by RNA editing. *Chem Rev* 106(8):3397-3411.
23. Goodman RA, Macbeth MR, & Beal PA (2012) ADAR proteins: structure and catalytic mechanism. *Curr Top Microbiol Immunol* 353:1-33.

APPENDIX A

PURIFICATION OF MUTANT PROTEINS

All mutant forms of human ADAR2 described in Chapter 2 were easily purified using the protocol described earlier. As an example, a Coomassie blue stained SDS-PAGE analysis of a mutant protein, E488Q, is shown (Fig. A1).

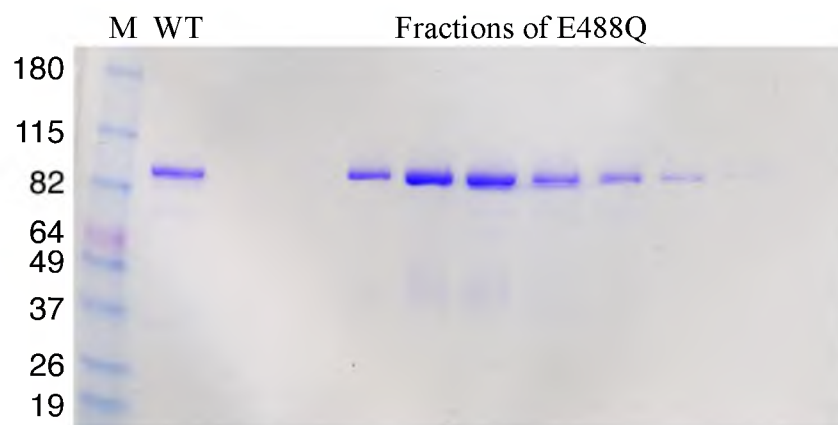


Figure A1. **Coomassie blue stained SDS-PAGE analysis of E488Q.** Fractions of E488Q mutant protein eluted from a Superdex 200 gel filtration column. Molecular weight standards (M) and WT hADAR2 (WT) as a reference are indicated.

APPENDIX B

HUMAN ADAR2 CONTAINING THE Y668H MUTATION

During our screen using the α -galactosidase reporter assay, we identified another mutant, Y668H, which edited the GAC hairpin-reporter more than WT hADAR2. This mutant exhibited low levels of editing of the GAC hairpin-reporter, with yeast colonies taking 2-5 weeks to turn faint green, but it edited the UAG hairpin-reporter well in the *in vivo* α -galactosidase reporter assay, similar to WT hADAR2.

As mentioned in Chapter 1, the deaminase domain of hADAR2 has an IP6 molecule in a cavity lined with basic and aromatic residues, and Y668 is one of these residues. The zinc ion in the active site of hADAR2 is coordinated by a His, two Cys and a water molecule. This is unlike most of the zinc containing hydrolases where the zinc is coordinated to two or three His. For example, ADA is coordinated to three His, an Asp and a water molecule. Among zinc coordinating residues, Cys donates the maximum negative charge to zinc, thereby decreasing its Lewis acidity. Thus, substituting His with Cys would decrease the positive charge on zinc, and hence increase the pKa of the zinc bound water. This would in turn decrease the catalytic efficiency of the reaction. Like ADARs, the zinc ion in the active site of CDA is also coordinated to one His, two Cys and a water molecule, yet this enzyme efficiently deaminates cytidine, which has been proposed to be due to compensating interactions. These compensating interactions could

be provided by the hydrogen bonds from one of the Cys thiolates (C129) to two backbone amide groups. However, the analogous Cys in ADAR2 (C451) is hydrogen bonded to only one backbone amide group. Thus, the second compensating interaction in ADARs could be provided by another zinc coordinating Cys, (C516), which is connected to IP6 by a relay of hydrogen bonds involving residues K483, D392 and K519. Thus, apart from playing a structural role, IP6 could also be required for catalysis by fine-tuning the environment of the active site zinc.

Since Y668 directly interacts with IP6, it is possible that the Tyr to His mutation at this position increases the positive charge on zinc, thereby decreasing the pKa of water coordinated to zinc, which will in turn increase the catalytic rate. If a mutation affects the charge on zinc, then it will result in a shift in the optimum pH for that enzyme. Therefore, the effect of the Y668H mutation on the charge of zinc could be tested by plotting k_{deam} as a function of pH. Potassium acetate buffer could be used to obtain a pH range of 3.8-5.4 and potassium phosphate buffer could be used to obtain a pH range of 5.8 to 7.5.

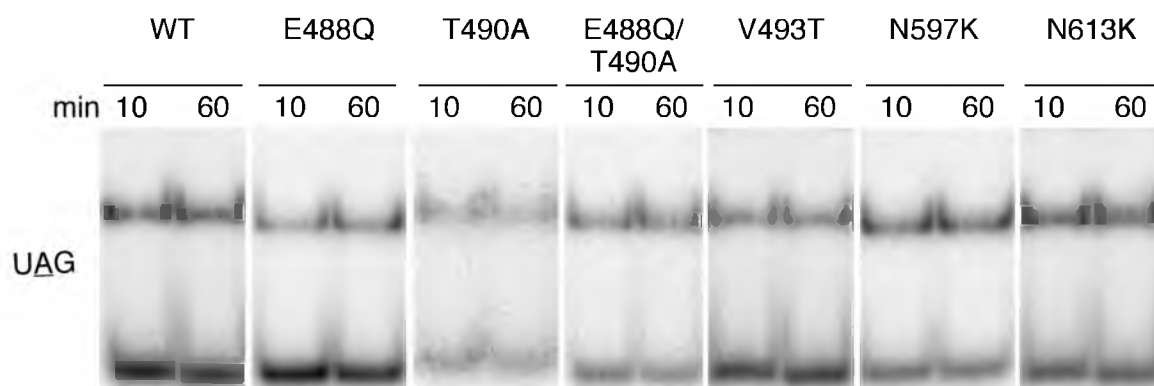
If the hypothesis that Y668H increases the positive charge on zinc is true then this mutant might slightly increase editing of both UAG and GAC. Since the α -galactosidase reporter assay is not sensitive enough to discern small differences in catalytic rates, Y668H must be purified and rates determined. We could not purify this mutant protein in two attempts using the previously described protocol. Since IP6 has a structural role and Y668 interacts directly with IP6, it is possible that mutating this residue decreases the stability of this protein. Thus, adding IP6 to the yeast media during protein expression might be helpful.

APPENDIX C

EQUILIBRIUM EXPERIMENTS

Equilibrium experiments were performed to determine the time for which the protein and RNA hairpin must be incubated before stopping the reaction by loading directly onto a 6 % (37.5:1 acrylamide/ bis-acrylamide) native gel running at 150V. The conditions used for the equilibrium experiments were identical to those used for the gel shift assays, except that incubations were done for 10 min and 60 min timepoints. Our experiments showed that the reactions were at equilibrium at 10 min, and further incubation for 60 min did not increase the bound fraction any further. For convenience, a 20 min timepoint was selected for the gel shift assays.

A



B

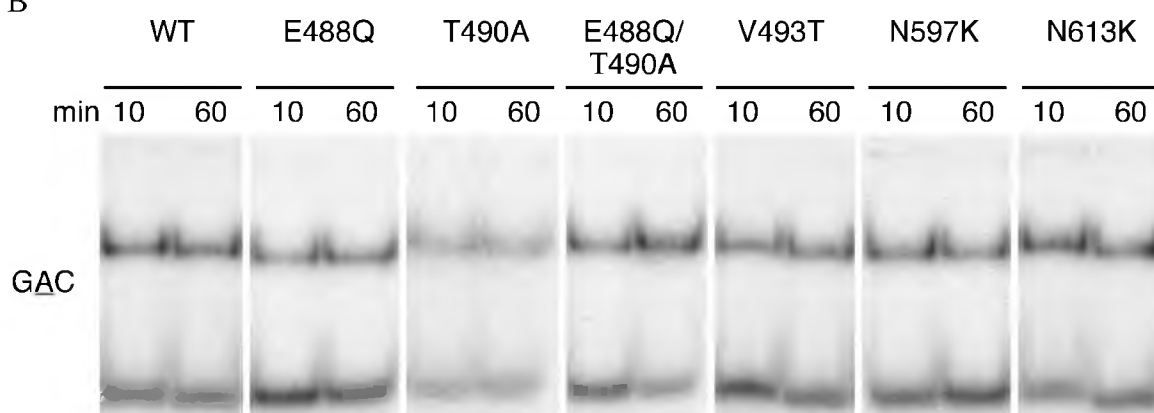


Figure C1. **Equilibrium experiments.** (A and B) Equilibrium experiments for WT hADAR2 and mutant proteins at 10 min and 60 min timepoints with UAG and GAC hairpins respectively.

APPENDIX D

INCUBATION TIMES REQUIRED FOR ACHIEVING ~20 %

EDITING OF A 418 bp dsRNA

An initial time course was performed as described (see Chapter 2, supplementary methods) with internally radiolabeled 418 bp dsRNA, to determine the time required to achieve ~20% overall editing. Once this time was determined, preference assays were performed for the same duration. The incubation time, incubation temperature and the average % editing for WT and mutant proteins in the preference assay are shown in Table D1.

Table D1. **Incubation time for achieving ~20 % editing**

Protein	Incubation time	Incubation temperature	Average % Edited
WT	~ 11 min	20 °C	19.9
E488Q	~ 6 min	20 °C	21.3
V493A	~ 12 min	20 °C	19.6
N597K	~ 16 min	20 °C	19.0
N613K	~ 13 min	20 °C	20.6
A589V	~ 14 min	30 °C	20.0
G336D	~ 17 min	30 °C	19.4
E488Q/T490A	~ 26 min	20 °C	25.1
WT	~ 5 min	30 °C	18.7
T490A	~ 56 min	30 °C	19.2

APPENDIX E

PREFERENCE ASSAYS OF A589V, G336D AND THE DOUBLE MUTANT E488Q/T490A

All preference assays were performed and analyzed as mentioned in Chapter 2 for WT hADAR2 at 20 °C, and are shown in Fig. S3. Surprisingly, A589V did not show an increase in editing of GAC, although it was selected for editing GAC in the α -galactosidase reporter assay. Rather, it showed a small decrease in editing of GAC. This could be because of many reasons: firstly, the α -galactosidase reporter assay was done using the UAG hairpin, whereas for the preference assay we used a 418 bp dsRNA. Therefore, it is possible that compared to WT hADAR2, this mutant increased editing of the UAG hairpin slightly, but it could not edit UAG in a perfectly base paired duplex as well. Secondly, even in the α -galactosidase reporter assay this mutant exhibited low levels of editing of the GAC hairpin-reporter, with yeast colonies taking 2-5 weeks to turn faint green. Therefore, it is possible that A589V was selected in the screen due to the fact that the screen is not as sensitive as the *in vitro* assays. A589V could have been selected also due to human error, since we did not use a colorimetric assay for detection of α -galactosidase, and the green yeast colonies were picked visually.

Overall A589V showed a small increase in specificity, $S_{rel} = 1.16$ (S_{rel} has been defined in Chapter 2) (Table S2); triplets poorly edited by WT hADAR2 were edited

almost to the same extent, however, triplets well edited by WT hADAR2 were edited to a greater percentage by the A589V mutant.

The G336D mutant, on the other hand, showed similar specificity as WT hADAR2, with $S_{\text{rel}} = 0.96$ (Fig. S3, Table S2); most of the triplets were edited to the same extent as observed with WT hADAR2. However, compared to WT hADAR2, G336D showed a slight increase in editing of GAC. In the case of the double mutant E488Q/T490A, most of the triplets were edited to a level intermediate between that observed with E488Q and T490A mutants alone, with $S_{\text{rel}} = 0.97$.

Table E1. Relative nearest neighbor specificity, S_{rel} , for mutants compared to WT hADAR2.

hADAR2 proteins	S_{rel}
WT	1
A589V	1.16
G336D	0.96
E488Q/T490A	0.97

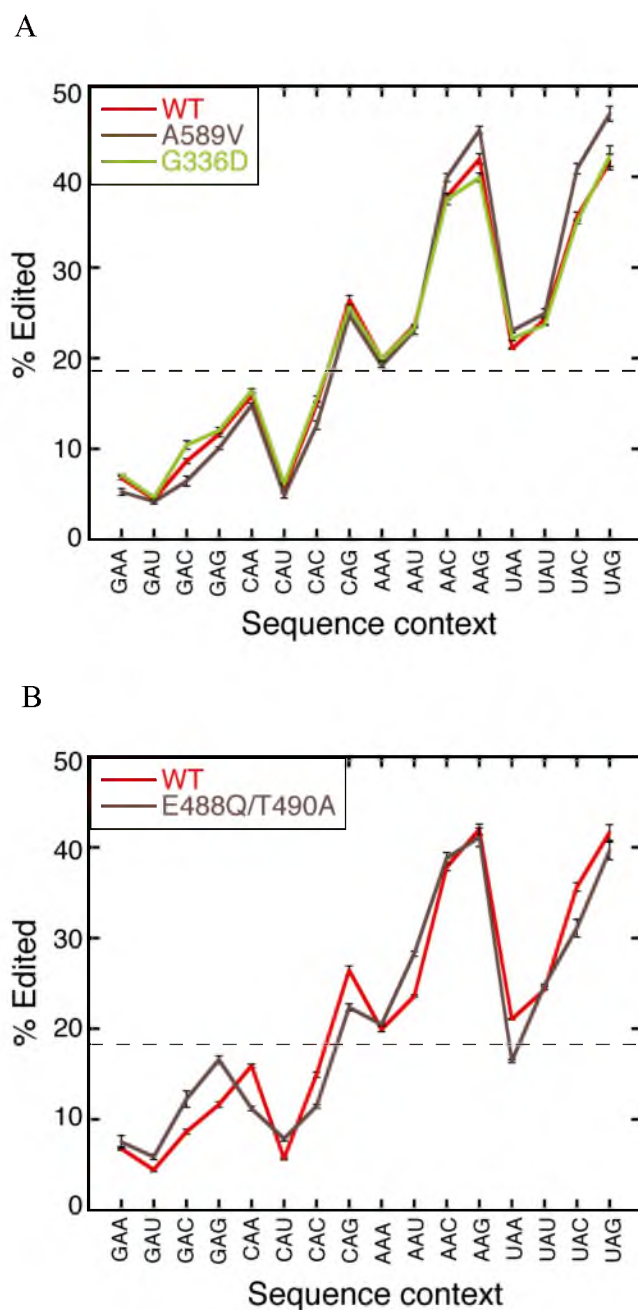


Figure E1. **Preference assays.** (A) Plot showing average % editing of adenosine in each of the 16 possible triplet contexts, determined from analysis of editing in a 418 bp dsRNA, for A589V and G336D compared to WT hADAR2. 196 adenosines were used to calculate average % editing of adenosine in 16 triplet contexts. Triplets are ordered in terms of 5' nearest neighbor, with 5' G followed by C, A and U. Mean % editing across 196 adenosines was 19.9% for WT hADAR2, 20.0% for A589V and 19.4% for G336D, and was normalized to 20% as indicated by the dotted line. Error bars = s.d. ($n \geq 3$). (B) As in (A), comparing E488Q/T490A to WT hADAR2. Mean % editing across 196 adenosines was 25.1%.

APPENDIX F

SEQUENCE OF 418 bp dsRNA USED FOR PREFERENCE ASSAYS

Sense strand:

5'-GTAGCCTTGAGCTTGGATCTGCCCAGCTTGGCGAGATTTTCAGGAGCTAAG
GAAGCTAAAATGGAGAAAAAATCACTGGATATACCACCGTTGATATATCCC
AATGGCATCGTAAAGAACATTTTGAGGCATTTTCAGTCAGTTGCTCAATGTACC
TATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGA
AAAATAAGCACAAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATG
AATGCTCATCCGGAATTCCGTATGGCAATGAAAGACGGTGAGCTGGTGATAT
GGGATAGTGTTACCCCTTGTTACACCGTTTTCCATGAGCAAACCTGAAACGTTT
TCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATA
TTCGCAAGATGTGGCGTGTTGCTTCCTAA-3'

Antisense strand:

5'-GATAGGCCAGGTTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTA
GAAACTGCCGGAAATCGTCGTGGTATTCCTCCAGAGCGATGAAAACGTTTC
AGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACC
AGCTCACCGTCTTTCATTGCCATACGGAATTCCGGATGAGCATTTCATCAGGCG
GGCAAGAATGTGAATAAAGGCCGGATAAACTTGTGCTTATTTTTCTTTACGG
TCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGA
GCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATC
AACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGA
AAATCTCGCCAAGCTGGGATCCAAACCTGAA-3'